DOCUMENT RESUME

ED 325 897 CS 507 338

AUTHOR	Studdert-Kennedy, Michael, Ed.
TITLE	Status Report on Speech Research. January-June 1990.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
REPORT NO	SR-101/102
PUB DATE	90
NOTE	298p.
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - General (020)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Articulation (Speech); *Language Acquisition; *Language Processing; Linguistics; Phoneme Grapheme Correspondence; *Phonology; Reading; *Speech Communication; Speech Habits; Vowels; Word Recognition
IDENTIFIERS	*Coarticulation (Speech); Speech Research; Voice Stress Analysis

ABSTRACT

This collection of articles is one of a series of semiannual reports on the status of speech research at Haskins' Laboratories. The titles of the 18 articles and their authors are as follows: "The Alphabetic Principle and Learning to Read" (Isabelle Y. Liberman and others); "Language Development from an Evolutionary Perspective" (Michael Studdert-Kennedy); "Phonemic Segmentation as Epiphenomenon: Evidence from the History of Alphabetic Writing" (Alice Faber); "Phonemic Similarity Effects and Prelexical Phonology" (Georgije Lukatela and M.T. Turvey); "Automatic and Prelexical Computation of Phonology in Visual Word Identification" (Lukatela and Turvey); "Phonological Ambiguity and Lexical Ambiguity: Effects on Visual and Auditory Word Recognition" (Ram Frost and others); "Morphological Relationships Revealed through the Repetition Priming Task" (Laurie Beth Feldman); "Listener-Talker Attunements in Speech" (Carol A. Fowler); "An Electropalatographic and Acoustic Study of Consonant-to-Vowel Coarticulation" (Daniel Recasens); "Coarticulation is Largely Planned" (D. H. Whalen); "The Perceptual Effects of Child-Adult Differences in Fricative-Vowel Coarticulation" (Susan Nittrouer and D. H. Whalen); "'Targetless' Schwa: An Articulatory Analysis" (Catherine P. Browman and Louis Goldstein); "Methods for Least-Squares Parameter Identification for Articulatory Movement and the Program PARFIT" (Richard S. McGowan and others); "'Clock' and 'Motor' Components in Absolute Coordination of Rhythmic Movements" (M. T. Turvey and others); "An Analogy between the Mucosal Waves of the Vocal Folds and the Wind Waves on Water" (Richard S. McGowan); "Laryngeal Control of the Plosives of Standard Thai" (Arthur S. Abramson); "Distinctive Vowel Length: Duration vs. Spectrum in Thai" (Arthur S. Abramson and Nianqi Ren); and "Effects of Preceding Context on the Voice-Onset-Time Category Boundary" (Bruno H. Repp and Hwei-Bing Lin). A list of DTIC and ERIC numbers is appended. (SR)

ED325897

Haskins Laboratories Status Report on
Speech Research

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Dadourian

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

SR-101/102
JANUARY-JUNE 1990

Haskins Laboratories Status Report on

Speech Research

SR-101/102
JANUARY-JUNE 1990

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Michael Studdert-Kennedy

Production Staff
Yvonne Manning
Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888
Contract NO1-HD-5-2910

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8820099

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594

Investigators

Arthur Abramson*
Peter J. Alfonso*
Thomas Baer*
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Lois G. Dreyer*
Alice Faber†
Laurie B. Feldman*
Janet Fodor*
Anne Fowler
Carol A. Fowler*
Louis Goldstein*
Carol Gracco†
Vincent Gracco
Vicki L. Hanson*
Katherine S. Harris*
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Koichi Tsunoda[1]
Michael T. Turvey*
Douglas Whalen

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Marion MacEachron*
Yvonne Manning
Joan Martinez
Lisa Mazur
Maura Murphy
William P. Scully
Richard S. Sharkany
Zefang Wang
Edward R. Wiley

Students*

Jennifer Aydelott
Dragana Barac
Paola Bellabarba
Melanie Campbell
Sandra Chiang
André Cooper
Margaret Hall Dunn
Terri Erwin
Elizabeth Goodell
Joseph Kalinowski
Deborah Kuglitsch
Simon Levy
Katrina Lukatela
Diana Matson
Gerald W. McRoberts
Pamela Mermelstein
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Nian-qi Ren
Christine Romano
Arlyne Russo
Richard C. Schmidt
Jeffrey Shaw
Caroline Smith
Jennifer Snyder
Robin Seider Story
Rosalind Thornton
Mark Tiede
Qi Wang
Yi Xu

*Part-time
†NRSA Training Fellow
[1] Visiting from University of Tokyo, Japan

Contents

The Alphabetic Principle and Learning to Read
Isabelle Y. Liberman, Donald Shankweiler, and Alvin M. Liberman 1

Language Development from an Evolutionary Perspective
Michael Studdert-Kennedy 14

Phonemic Segmentation as Epiphenomenon: Evidence From the History of Alphabetic Writing
Alice Faber 28

Phonemic Similarity Effects and Prelexical Phonology
Georgije Lukatela and M. T. Turvey 41

Automatic and Prelexical Computation of Phonology in Visual Word Identification
Georgije Lukatela and M. T. Turvey 74

Phonological Ambiguity and Lexical Ambiguity: Effects on Visual and Auditory Word Recognition
Ram Frost, Laurie B. Feldman, and Leonard Katz 85

Morphological Relationships Revealed Through the Repetition Priming Task
Laurie Beth Feldman 101

Listener-talker Attunements in Speech
Carol A. Fowler 110

An Electropalatographic and Acoustic Study of Consonant-to-Vowel Coarticulation
Daniel Recasens 130

Coarticulation Is Largely Planned
D. H. Whalen 149

The Perceptual Effects of Child-Adult Differences in Fricative-Vowel Coarticulation
Susan Nittrouer and D. H. Whalen 177

"Targetless" Schwa: An Articulatory Analysis
Catherine P. Browman and Louis Goldstein 194

Methods for Least-squares Parameter Identification for Articulatory Movement and the Program PARFIT
Richard S. McGowan, Caroline L. Smith, Catherine P. Browman, and Bruce A. Kay 220

"Clock" and "Motor" Components in Absolute Coordination of Rhythmic Movements
M. T. Turvey, R. C. Schmidt, and Lawrence D. Rosenblum 231

An Analogy between the Mucosal Waves of the Vocal Folds and Wind Waves on Water
Richard S. McGowan 243

Laryngeal Control in the Plosives of Standard Thai
Arthur S. Abramson 250

Distinctive Vowel Length: Duration vs. Spectrum in Thai
Arthur S. Abramson and Nianqi Ren 256

Effects of Preceding Context on the Voice-Onset-Time Category Boundary
Bruno H. Repp and Hwei-Bing Lin 269

Appendix 289

Haskins
Laboratories
Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 1-13

The Alphabetic Principle and Learning to Read*

Isabelle Y. Liberman,† Donald Shankweiler,† and Alvin M. Liberman††

Proper application of the alphabetic principle rests on an awareness of the internal phonological (and morphophonological) structure of words that the alphabet represents. Unfortunately for the would-be reader-writer, such awareness is not an automatic consequence of speaking a language, because the biological specialization for speech manages the production and perception of these structures below the level of consciousness. Not surprisingly, then, awareness of phonological structures is normally lacking in preliterate children and adults; the degree to which it does exist is the best single predictor of success in learning to read; lack of awareness usually yields to appropriate instruction; and such instruction makes for better readers. That some children have particular difficulty in developing phonological awareness (and in learning to read) is apparently to be attributed to a general deficiency in the phonological component of their natural capacity for language. Thus, these children are also relatively poor in short-term memory for verbal information, in perceiving speech in noise, in producing complex speech patterns, and in finding the words that name objects. All children will benefit from instruction that is intelligently designed to show them what the alphabet is about.

For some 15 years, we have been exploring the sources of the problems beginners encounter in learning to read. Since children are quite fluent in their native language when first encountering the language in print, we began by asking what seemed to us the obvious question: what is required of the child in reading a language but not in speaking or listening to it? The first answer that came to mind, of course, was the discrimination of the visual shapes of the letters. But investigators who had done comprehensive studies of many different aspects of the reading process (see Doehring, 1968), or who had compiled exhaustive reviews of the visual factors involved in reading (Benton & Pearl, 1978; Stanovich, 1982; Vellutino, 1979; Vernon, 1957), were all in agreement that beginners who were making little progress in learning to read generally showed no significant difficulty in the visual identification of letters.

Beyond letter identification, reading requires mastery of a system that maps letter shapes to units of speech. However, as we noted years ago (Liberman, 1973), there is no evidence that children of normal intelligence, given proper instruction, have difficulty in associating individual letters of the alphabet with their appropriate speech equivalents. Perhaps, then, they are defeated by the often complex and irregular relations in English between spelling and language. Surely, the complexities of English spelling do create some problems. But even when the items to be read include only those words that map the sound in a simple, consistent way, many children still fail (Savin, 1972).

Learning to identify the letters, learning to associate them with consonant and vowel sounds, learning to cope with the irregularities of English spelling—none of these is the primary obstacle in learning to read. What is it then that makes reading so hard while speech is relatively so easy? In the seventies, we (Liberman, 1971; Shankweiler & Liberman, 1972) and other investigators (Elkonin, 1973; Gleitman & Rozin, 1977; Klima, 1972; Mattingly, 1972) proposed another possible source of difficulty in reading that is not present in

---

Parts of this paper were adapted from "Phonology and the problems of learning to read and write." *Remedial and Special Education*, 1985, *6*, 8-17. This research was supported in part by grant HD-01994 to Haskins Laboratories and by grant NIH-21888 to Yale University/Haskins Laboratories from the National Institute of Child Health and Human Development.

speech. Although both reading and speech require some degree of mastery of language, reading requires, in addition, a mastery of the alphabetic principle. This entails an awareness of the internal phonological structure of the words of the language, an awareness that must be more explicit than is ever demanded in the ordinary course of listening and responding to speech. If this is so, it should follow that beginning learners with a weakness in phonological awareness would be at risk.

We here first set forth the considerations that led us to that view, followed by the evidence that supports it. Then we say why we should consider that deficits in awareness of the phonological structure may be only one symptom of a more general underlying deficiency in the phonological component of the beginning reader's capacity for language. Finally we consider the implications for instruction.

PHONOLOGY AND THE ALPHABETIC PRINCIPLE

We begin, then, with the assumption that reading by an alphabetic writing system requires mastery of the alphabetic principle. Surprisingly, this assumption, which seems to us a truism, is not accepted by everyone in the field, as we will see. But even among those who think the principle important, many take it—we would say mistake it—to mean simply an understanding by the would-be reader that the discrete letters of the alphabet represent the discrete sounds of speech. Our view (Liberman, 1983) is different. As we see it, the letters of the alphabet do not represent sounds as such, but rather the more remote phonological (and morphophonological) segments those sounds convey. This is not to quibble. For surely it must be somewhat confusing to children to be told that the word *bag* is spelled with three letters, when his ears tell him plainly that it has but one sound. The confusion is only worse confounded if the teacher insists, against the evidence of what the child hears, that *bag* can be divided into three sounds, and that these can then be "blended" so as to re-form the word. For there is, in fact, no way, with or without the marvels of modern technology, to divide *bag* into pieces of sound that correspond in any reasonable way to the sounds of the three letters, nor is there any way to synthesize the word by somehow putting the letter sounds together. Though bag does truly consist of three segments—it differs from *sag* in the first, from *big* in the second, and from *bat* in the third—these segments are to be found only in the underlying phonology, not in the surface appearances of the sound (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

To emphasize that letters stand for sounds also risks making it that much harder for a reader to understand perfectly reasonable aspects of English spelling—for example, *cats* and *dogs*, instead of *cats* and *dogz*; or *bat* and *batter*, instead of *bat* and *badder*; or *an apple*, instead of *uh napple*; and so on. As for those aspects of English spelling that are most egregiously unreasonable—for example, *through*, *rough*, and the like—we confess that even a proper understanding of the alphabetic principle is not likely to be of much help, but then neither will anything else, short of learning about how the language has changed since the orthography was developed. (Such spellings need not be a hindrance if they are introduced only after the more systematic aspects of the orthography have been understood.)

The identification of letters with sounds promotes yet another misunderstanding, this one about the nature of words and how they are perceived. For it accords all too well with the commonplace notion that it is only spoken words that are made of phonological units. Words that come to us via print are incorrectly thought to be different, in that they supposedly can (and perhaps should) be perceived independently of phonology. On this basis, some advise that the reader be taught from the very beginning to skip the phonology (read this as skipping the sound) and go "direct to meaning" (Goodman, 1976; Smith, 1971). Others grant that going through the phonology—which is taken to mean "sounding it out"—may be useful for the beginning reader, or for the mature reader who encounters a strange word, but they otherwise hold that the putatively "direct" (nonphonological) route is the way to go (Coltheart, 1978; Waters, Seidenberg, & Bruck, 1984).

We believe that these assumptions seriously misconstrue the nature of words and the processes by which they are produced and perceived, in print as in speech. Consider, in this connection, a critical difference between language and all other natural forms of communication. In all the nonlinguistic systems—whether the medium is acoustic, optical, electrical, or chemical—meanings are conveyed by signals that differ holistically, one from another. This is to say that there are no words. The inevitable consequence is that the number of meanings that can be communicated is limited to the number of holistically different signals the animal can produce and perceive, a number that is always quite small. Even if that number can somehow be increased, there is no way of doing it so as to guarantee that the new signals will be immediately recognized as belonging naturally to a system that has a specifically communicative function.

Language is different in a most important way. Meanings are not conveyed directly by signals that differ holistically, but rather by words that are distinct from each other in their internal structure. This structure is formed of a small number of meaningless phonological segments we know as consonants and vowels, and governed according to a highly systematic combinatorial scheme called phonology. The consequence is that words can (and do) number in the tens of thousands. Moreover, there is a perfectly natural basis for accommodating new words, since the phonological system, which all speakers of the language have in common, automatically recognizes a new, but legal, structure as a word that stands ready to have meaning attached to it. It is only because children have this phonological system that they are able to acquire new words and their meanings with such astonishing ease and rapidity (Studdert-Kennedy, 1987).

What follows, then, is that phonology governs all words, whether dead, living, or waiting to be born. So, whatever else a word is, and regardless of whether it is spoken or printed, it is always a phonological structure. If listeners or readers correctly perceive a word, they correctly perceive the structure that distinguishes it. They may very well be unsure of its meaning—indeed, they may even have got the meaning wrong—but if they have the phonological structure, they have a perfectly adequate basis for ultimately getting its meaning properly sorted out. As for going directly to meaning—that is, independently of phonology—surely that is done when a person sees a picture, for example, or hears the roar of a lion, but not when one perceives a word as it is spoken or read.

From our point of view, then, there is no reason to ask, as some do, whether readers must, or should get to meaning via the phonology. To make sense of this question, one must make three false assumptions. The first is that the meanings can be communicated in language independently of words (that is, phonological structures), but in fact they cannot. The second is that the phonological units that form all words are equivalent to the sounds of speech, but in fact they are not. And the third is that an alphabetic transcription specifies, on a segment by segment basis, how the speech organs are to be articulated and coarticulated so as to produce the sounds of speech, but in fact it does not. What the reader must do is to match the alphabetic transcription to the abstract phonological structure of the word it represents. In the case of a familiar word, this structure and its associated meaning(s) are available in the reader's lexicon; in the case of an unfamiliar word, given a command of the alphabetic principle, the structure is easily formed and thus made ready for whatever meanings may subsequently be attached to it. Once the reader has the phonological form of the word, the appropriate phonetic structure and its associated articulatory movements are automatically available to him for use in working memory, or for reading aloud if the occasion should call for that.

PHONOLOGICAL PROCESSING IN READING-WRITING AND IN LISTENING-SPEAKING

Why is it normally so much harder and less natural to deal with phonological structures in reading and writing than it is in listening and speaking? A serious attempt to answer this question would take us quite deeply into the phonological system and its biology, for the answer requires explaining, among other things, why speech could have evolved in the history of our species but writing systems could not, and why speech can develop in the child without explicit instruction but reading and writing typically can not (Liberman, 1989). Here, we can offer only a truncated account.

Like all members of the animal kingdom, human beings have highly specialized ways of communicating with their fellows. In the human case, and only there, this specialization includes, as a critical component, the phonological system that, as we have seen, makes large vocabularies possible. As this system evolved in the race, and as it develops anew in each child, it employs abstract motor structures—let us call them gestures—that ultimately control the movements of the speech organs (Browman & Goldstein, 1985; Liberman & Mattingly, 1985; Liberman & Mattingly, 1989). These gestures are adapted for one purpose and for one purpose only: the production of strings of consonants and vowels at rates many times more rapid than could otherwise be achieved. These rates, which run about eight to ten per second on average, are managed by precisely overlapping and merging the articulatory movements that produce the phonologically significant aspects of the speech sounds. This coarticulation, as it is called, is a most complex process, but it does not appear so in the normal exercise of speech functions, because it is done automatically and naturally by this aspect of the specialization for language. A consequence is that a neurologically normal child, put in a speech environment, can hardly be prevented from learning to form phonological structures and to exploit coarticulation for that purpose. (Lacking this specialization, nonhuman primates do not, and cannot, learn to produce these structures; this is to say that they cannot produce words.) A more important consequence for our purposes is that to speak a word one need not know how it is spelled. The speaker need only think of the word; the phonological component of his grammar "spells" it for him. Indeed, the automaticity of this specialization makes it that much harder to be aware of how the word is spelled, or even to know that such a thing as spelling exists.

Perception of the speech signal is correspondingly complex and automatic. Given coarticulation, there is no direct correspondence between the phonological structure intended by the speaker and the surface properties of the sound. Most relevant to our concerns is the fact that, as we have so often pointed out, the number of segments in the sound is not equal to the number of segments in the phonological structure it conveys (Liberman et al., 1967). Thus, the three consonants and vowels of a word like *bag* are so thoroughly coarticulated as to produce a single segment of sound. But this is no problem for listeners, for they have only to rely on their phonological specialization to automatically process the speech signal and recover the coarticulated gestures that caused it (Liberman & Mattingly, 1989). It is a problem for would-be readers, however, because, given the complex relation between phonological structure and sound, and the automaticity with which this relation is dealt with in speech, they find it just that much harder to be aware that the word does have an internal structure and thus to appreciate why an alphabetic transcription makes sense.

Small wonder then that an alphabetic writing system is such a comparatively recent development in the history of our species. In contrast to the naturally evolved phonological structures it represents, it is an artifact. The development of this artifact had to wait on the discovery—and it was a discovery—that words have an internal structure. Once that discovery was available, someone could and did invent the notion that, by representing the units of that structure with arbitrarily chosen optical shapes, people could read and write all the words of the language—those they were already familiar with and those they had yet to encounter. But they could exploit this wonderful invention only if they understood the discovery on which it was based.

AWARENESS OF PHONOLOGICAL STRUCTURE AND READING

Development of phonological awareness in children

Considerations such as these led us at the very outset of our research on reading to suppose that preliterate children would not naturally have made the discovery that underlay the invention of the alphabet, from which it would follow that they would not be prepared to understand and apply the alphabetic principle. So we began to examine developmental trends in phonological awareness by testing the ability of young children to segment words into their constituent elements. We investigated the children's segmentation of spoken words both by syllable and phoneme (Liberman, Shankweiler, Fischer, & Carter, 1974). (The latter class of units of the phonological representation comprise the consonants and vowels. Heretofore we have referred to these only by the general term, *phonological segments*. From now on we will call then *phonemes* to distinguish them from syllables). We found that normal preschool children performed rather poorly, but that the phonemes presented the greater difficulty by far.

It was clear from these results that awareness of phoneme segments, the basic units of the alphabetic orthography, is initially harder to achieve than awareness of syllable segments, and develops later, if at all. More relevant to our present purposes, it was also apparent that a large number of children, about 30 percent of our sample, had not attained an understanding of the internal phonemic structure of words, even at the end of a full year in school. Surely, they are the ones we need to worry about, because they are the ones who are deficient in the linguistic awareness that may provide entry into the alphabetic system.

Lack of phonological awareness and reading failure in children

Is lack of phonological awareness in fact related to failure in reading and writing? That the answer is yes is strongly supported by studies in a number of languages. In English, the relation has been found, for example, in studies by Blachman (1984); Bradley and Bryant (1983); Fox and Routh (1980); Goldstein (1976); Helfgott (1976); Treiman and Baron (1981); Vellutino and Scanlon, (1987). Their findings have been supported by studies in Swedish by Lundberg and associates (Lundberg, Olofsson, & Wall, 1980) and Magnusson and Naucler (1987), in Spanish by de Manrique and Gramigna (1984), in French by a group of Belgian researchers (Alegria, Pignot, & Morais, 1982), and recently in Italian by Cossu and associates (Cossu, Shankweiler, Liberman, Tola, & Katz, 1988).

The study carried out by Lundberg and his associates in Sweden (Lundberg et al., 1980) is worthy of special mention on two counts. It provides one of the most intensive examinations of the linguistic abilities of the kindergartners. It is noteworthy as well because it also addresses the question of whether the children's deficiency is, in fact, linguistic or whether it might be attributable to a deficiency in general analytic ability. Their battery of 11 tests given to 200 kindergartners included both linguistic and nonlinguistic tasks. In the linguistic set were: (1) word synthesis tasks that varied in two dimensions of two levels each—with or without memory load and using either phoneme or syllable units, and (2) word analysis tasks analogous to those for synthesis and, in addition, three others demanding analysis of phoneme position in words, reversal of phoneme segments in words, and rhyming. Since the linguistic tasks required the child to shift attention from meaning to abstract form, thus possibly reflecting a general cognitive function not exclusively limited to linguistic material, nonlinguistic control tasks that simulated those cognitive demands were also included. The most powerful predictors of later reading and writing skills in the entire battery turned out to be those requiring phonological awareness, specifically the analytic ability to manipulate phonemes in words. In contrast, the poor readers showed no particular deficiency in the non-linguistic tasks.

These findings that tasks of linguistic, rather than nonlinguistic analysis, and specifically phonemic analysis, were predictive of reading failure have since received support from other studies here and abroad. For example, in a study of six to nine-year-olds with severe reading disability (Morais, Cluytens, & Alegria, 1984) it was found that these children were poorer on segmenting words into their constituent parts but performed just as well as normal readers in a matched task that required them to deal analytically with musical tone sequences instead of words. The question of a possible general analytic deficit was also addressed in two complementary experiments, one with good and poor readers in the third grade and the other with good and poor readers in adult education classes (Pratt, 1985). All the subjects were given three linguistic awareness tests and one nonspeech control task identical in format to one of the linguistic measures. Significant differences were found between the good and poor readers at both age levels on all three linguistic measures, but not on the nonspeech control task. Thus the poor readers, whether young or old, had no more difficulty in segmental analysis than the good readers when the task was nonlinguistic; their problem was limited to the segmental analysis of speech.

Not only reading, but early spelling proficiency has also been found to be closely related to analytic phonological skills. In a study of kindergartners (Liberman, Rubin, Duques, & Carlisle, 1985), the children's ability to produce invented spellings, given only some knowledge of letter names (see Read, 1986), was related to their performance on a series of language-based tasks. It was found that the children's proficiency in spelling was more closely tied to phonological awareness than to any of the other aspects of language development tested. Of the eight tasks in the study, only the three which unquestionably tapped phonological analysis skills made a difference statistically. They combined to account for 93 percent of the variance in proficiency in invented spelling. A phoneme analysis test patterned after Lundberg et al. (1980) made the largest contribution—67 percent of the variance. A test of the ability to write letters to phoneme dictation accounted for 20 percent more, and a phoneme deletion task ("Say *milk* without the *m*") added another 6 percent. (A fourth task, picture naming, added 1 percent, but did not reach significance in the correlation. As we will note later, naming can, however, be a subtle indicator of more general phonological difficulties.)

Among the four language-based tests that did not contribute to the invented spelling performance were three that are frequently included in clinical evaluations—receptive vocabulary, articulation as measured by the repetition of simple words, and letter naming or writing. The fourth was a syllable deletion test ("Say *bookcase* without the *book*"). Being able to segment words by syllable was, as we would expect, not enough to equip the child to produce alphabetically written words.

The remedial effect of training in awareness

Given the abundant evidence that phonological awareness is predictive of success in reading, it is of interest to know that such awareness can be trained even in preschool and kindergarten (Bradley & Bryant, 1983; Content, Morais, Alegria, & Bertelson, 1982; Lundberg, Frost, & Peterson, 1988; Olofsson & Lundberg, 1983). It is of special interest to find, moreover, that the training can have a salutary effect on future reading skill. Impressive evidence for the efficacy of the training comes from a pair of experiments by Bradley and Bryant (1983). The first experiment confirmed the high correlations found by others between preschoolers' phonological awareness and later reading skill. This was done by a comparison of children's performance on rhyming tasks and their achievement in reading and spelling several years later. The second experiment was directed to an examination of the effect of various kinds of early training on the later academic achievements of children considered to be at risk for failure. To this end, the children who had been found in the first experiment to have a low level of phonological awareness were divided into four groups. One group was trained to sort pictured words on cards by phonological categories. A second received the same training except that letters corresponding to the phonemic categories were added. A third group was trained to sort by semantic categories. A control group was given equal time and unrelated card play. The two phonologically trained groups were found to be superior to the others in subsequent tests of reading and spelling. Moreover, in follow-up studies, they continued to maintain their advantage.

Further evidence for the positive effect of early training in phonological awareness is found in a longitudinal study by Lundberg and associates (Lundberg, et al., 1988). An experimental group of kindergartners who had participated in a variety of analytic word games was found at year-end to be superior in phoneme awareness to a matched control group. When compared in academic achievement in the first grade, the experimental group was slightly below the controls in math and IQ but significantly superior in both reading and spelling. Moreover their advantage was maintained when the children were retested in the second grade.

Lack of phonological awareness and adult literacy

What about phonological awareness in adult nonreaders? Is it still a problem for them? The question as to whether phonological awareness improves spontaneously with age or requires some form of instruction is a crucial one, with obvious implications not only for preschool instruction but also for the design of literacy teaching programs geared to adolescents and adults. This question was explored in an unusual investigation by a Belgian research group who examined the phonological awareness of illiterate adults in a rural area of Portugal (Morais, Cary, Alegria, & Bertelson, 1979). They found that the illiterate adults could neither delete nor add phonemes at the beginning of nonsense words, whereas others from the same community who had received reading instruction in an adult literacy class succeeded in performing those tasks. The authors concluded that awareness of phoneme segmentation does not develop spontaneously even by adulthood but arises as a concomitant of reading instruction and experience.

In view of these findings, we believed it would prove of value to test the phonological awareness of adults who had had reading instruction but were nonetheless poor readers. To this end, our reading research group (Liberman, et al., 1985) tested the members of a community literacy class, all of whom were having serious decoding problems despite years of schooling. What we found was that these adults performed with difficulty on a very simple task in which the subjects were required only to identify the initial, medial, or final sound in monosyllabic words. Though this is an exercise that one might expect a first grader to be able to perform, our adults managed to produce correct responses on only 58 percent of the items. Moreover, they clearly found it to be singularly frustrating and unpleasant. This inability of adults with literacy problems to perform well on tasks demanding explicit understanding of phonological structure has also been found by other investigators—Byrne and Ledez (1983) in Australia; Marcel (1980) in England; and Read and Ruyter (1985) in a prison population in the U.S.

A BROADER PHONOLOGICAL DEFICIENCY AND READING

Why do some people have difficulty in achieving the understanding of phonological structure that application of the alphabetic principle requires? One possibility, as we noted before, is that they may suffer from a general deficiency in the ability to divide objects of all kinds into their constituent elements. But as we pointed out, the results of several studies suggest that the difficulty is specifically linguistic. Another possibility—one that we and others have pursued—is that the poor reader's difficulty with analyzing words into their constituent units is one among several symptoms of a general deficiency in the phonological component of the child's natural capacity for language. If the underlying biology tends to set up phonological structures weakly, then it should follow that these structures would be that much harder for the child to bring to a level of explicit awareness. But there would be other consequences for the processing of language, and these we consider below.

Problems in short-term memory and sentence comprehension

Because short-term memory depends on the ability to gain access to phonological structure and to use it to hold linguistic information (Conrad, 1964; Liberman, Mattingly, & Turvey, 1972), we might expect people who have underlying phonological deficiencies to show various limitations on verbal tasks that tap short-term memory. This expectation is amply borne out.

The research literature contains many reports that young children who are poor readers are deficient in short-term memory. Typically they retain fewer items from a set of fixed size than age-matched good readers (see Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979; Shankweiler, Smith, & Mann, 1984; Wagner & Torgesen, 1987). However, memory difficulties for poor readers appear to arise only under specific conditions; chiefly, they occur when the items to be retained are words and nameable objects. When the test materials do not lend themselves to verbal (i.e., phonological) encoding, as in memory for nonsense shapes or unfamiliar faces, memory testing does not find poor readers at a disadvantage (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982). The problem seems, therefore, to be a material-specific one, not an all-embracing memory impairment.

It is noteworthy, in addition, that memory differences between good and poor readers may also depend on other demands of the task—tasks that require rote recall of a list of unstructured items may be less differentiating than tasks that require both storage and further processing of the incoming material as in sentence processing (Daneman & Carpenter, 1980; Perfetti & Goldman, 1976). Since language structures are hierarchically organized and sequentially transmitted, comprehension of language, either by ear or by eye, depends on a short-term memory system that transiently stores and continuously processes the incoming segments of the linguistic message. In keeping with current usage (see Baddeley & Hitch, 1974; also Shankweiler & Crain, 1986), we call this form of memory *working memory*, a term which is used throughout this book. A phonological deficiency would understandably impair the functions of working memory and could be expected, in turn, to have repercussions on comprehension, whether of spoken discourse or printed text. For example, in sentence processing, the parsing of phonological segments into lexical units and the grouping of these units into higher-level phrasal structures requires phased control of the flow of linguistic information through the language apparatus.

We could therefore expect that children with reading disability would sometimes comprehend sentences poorly because of their difficulties in setting up and retaining phonological structures. The difficulty should be especially acute in reading, where the problem of decoding from print would create an additional processing load in an unskilled reader who decodes poorly. The important insight that the lower-level and higher-level reading problems of the poor reader are causally connected through constriction of working memory was contributed by Perfetti and Lesgold (1979). In their terms, poor decoding skills coupled with the limitations of working memory create in the poor reader a "bottleneck" in information flow with severe repercussions for comprehension.

But, as Crain and Shankweiler discuss (Crain, 1989; Shankweiler, 1989), comprehension difficulties of poor readers are not limited to reading. From our working memory perspective, difficulties should also arise in spoken language processing, especially if the sentence material contains remote dependencies or structural ambiguities that necessitate reanalysis, or if the comprehension task presents additional complexities that further dilute memory resources. Several reports in the literature indicate that disabled readers do have problems in comprehending such sentences in spoken form as well as in reading (Byrne, 1981; Mann, Shankweiler, & Smith, 1984; Stein, Cairns, & Zurif, 1984). More recent findings (Shankweiler, 1989) indicate that the poor readers fail not because they lag behind their good reading peers in comprehension of grammar as such, but because working memory is overloaded due to deficient phonological processing.

By changing the task in various ways to reduce the demands on memory while testing the same grammatical structures, it has been shown that poor readers can succeed as well as good readers in comprehending complex grammatical structures (Crain, Shankweiler, Macaruso, & Bar-Shalom, 1990; Smith, Macaruso, Shankweiler, & Crain, 1989). Thus, a memory impairment stemming from a weakness in phonological processing can masquerade as a grammatical or semantic deficit.

Other language-related problems

Thus far we have discussed difficulties involving the phonological components of language that directly affect reading. Reading is affected both by the difficulties of accessing and mentally manipulating phonemic segments and by the limitations on use of the working memory that we have just discussed (though in the case of working memory the consequences are not confined to reading). We now turn to other deficits displayed by poor readers that are phonological in nature, but do not affect reading directly. These are worth mentioning, both for their diagnostic value, and because they add to the weight of the evidence that all the elements of the syndrome of many poor readers may stem ultimately from a deficiency in phonological processing.

One such deficit is suggested by some preliminary research into the speech perception of poor readers that was carried out by Brady and associates at Haskins Laboratories (Brady, Shankweiler, & Mann, 1983). In their experiments, good and poor readers were tested on two auditory perception tasks, one involving words and the other nonspeech environmental sounds. The identification tasks were presented under two conditions—with favorable and unfavorable noise ratios. The findings were that the poor readers did show a deficit, but it was specific to the speech stimuli and occurred only in the noise-masked condition. They did not differ from the good readers in the perception of nonspeech environmental sounds, whether the sounds were noise-masked or not. Note that the poor readers apparently needed a higher quality of signal than the good readers for error-free performance in speech, but not for nonspeech environmental sounds. These results suggest that the minor deficit displayed by the poor readers may derive from phonological structures that are set up more weakly than in good readers, or are more difficult to activate.

Additional evidence for a broader phonological deficit in poor readers is provided by a study of speech production, specifically, the errors of junior high school students (Catts, 1986). The critical finding was that the reading disabled students made significantly more errors than matched normals on three different tasks in which their speech production was stressed. The author concluded, as we would, that their difficulties in speech production may be an extension of deficits in the phonological realm.

More evidence for a broad phonological deficit in poor readers was provided by a study of the performance of second graders on a naming test (Katz, 1986). This study confirmed what others had shown—that poor readers named more words incorrectly. But it went further to show that their difficulties are often phonological and not semantic, as might be assumed. Three kinds of evidence were presented in this regard. first, when quizzed about the characteristics of the object they had named incorrectly (e.g., "tornado" for volcano), the poor readers were often able to describe it accurately. They clearly knew what the object was. That is, they described a volcano, not a tornado. Second, given the name of the item, they could select it from a group of pictured objects. That is, they could identify it correctly. And third, their naming errors were often related to the phonological and not the semantic aspect of the word. For example, though the name given to the picture of a volcano was incorrect, it shared syllable count, stress pattern, and vowels with the target word.

Distorted production of the word for an item that had been correctly identified could stem either from deficient specification of the phonological structure in the lexicon, or from deficient retrieval and processing of the stored phonological information. In either case, the source of the difficulty relates to the phonological structure of the words and not to their meanings.

PHONOLOGY AND THE SUCCESSFUL DEAF READER

The congenitally deaf constitute a population with a phonological impairment arising from an entirely different source. Surprisingly, this group represents some of the most compelling evidence for the importance of phonological abilities for reading. It is well known that profound deafness from birth or early life usually results in attainment of a low level of reading skill. The hearing impaired of all ages tend to read far below grade expectations. But, nonetheless, differences in reading achievement are related to differences in phonological abilities even in deaf populations. Moreover, a few congenitally, profoundly deaf individuals can read well, even up to the college level.

How are these successful deaf readers different from the majority? Vicki Hanson and her associates at Haskins Laboratories asked that question in a series of experiments (Hanson, 1982; Hanson & Fowler, 1987; Hanson, Liberman, & Shankweiler, 1984) reported on by Hanson (1989). Briefly, they found that the successful deaf readers were not limited to reading English words as if they were logographs; that is, they were not, as one might assume, dependent on a limited store of words learned in paired associate fashion as visual designs. The results, which are discussed in detail by Hanson (1989) showed that unlike their poor reading peers among the deaf, these subjects, despite so little exposure to sound, were able to access that knowledge both in reading and in retaining verbal material in short-term memory.

In reading, the good readers among the deaf displayed their phonological sensitivity by responding differentially to rhyming and nonrhyming pairs of words (*save/wave* vs. *have/cave*) and by being able to name the real word equivalents of nonwords (*flame* for f-l-a-i-m; *tall* for t-a-u-l). In a short-term memory experiment, the successful deaf readers were more affected by phonetically confusing words than by those that were orthographically confusing or whose signs were formationally confusing. These results certainly suggest that successful deaf readers are using phonological processing, a conclusion also reached by Conrad with a less severely impaired population of deaf readers (Conrad, 1979).

The question of how the congenitally, profoundly deaf might develop phonological sensitivity without being able to hear the sounds of speech is explored by Hanson (1989). She identifies several sources of information that may be helpful. The orthography itself tells them something about the systematic phonological forms of words. In addition, oral training when available supplies information about the gestures used to produce speech. Lipreading also provides considerable useable information, and the deaf individual's own attempts at speech may reflect more phonological sensitivity than is apparent to the hearing listener.

IMPLICATIONS FOR INSTRUCTION

In view of all the evidence that has accumulated in the past 15 years to support the critical importance of phonological sensitivity for the attainment of literacy in an alphabetic system, one would surely expect teacher training to reflect these findings. Unfortunately, all too often it does not. Many teachers are being trained to teach reading without themselves ever having learned how an alphabetic orthography represents the language, why it is important for beginners to understand how the internal phonological structure of words relates to the orthography, or why it is hard for children to achieve this understanding.

In fact, teachers are all too often being provided with an instructional procedure that directs them specifically not to trouble the learner with details of how the alphabet works. Instead, they are told to view reading as a "guessing game" (Goodman, 1976) in which the general import of the message, and not the actual words of the text, is to be emphasized. Beginners are encouraged to memorize the appearance of words as visual patterns by whatever means they can muster and to use their store of memorized words and their "whole language" capability as a basis for guessing the rest of the message from picture cues and context. Thus, they are not to be corrected when reading "kids" for *children* in a story about a playground, "Crest" for *toothpaste* in a story about dental hygiene, and "cats" for *dogs* in a story about pets.

Fortunately, many children—the lucky 75 percent or so who learn to read whatever the method—manage to pick up the alphabetic principle without much explicit instruction, if any. That is, given experience with printed material, they begin to discover for themselves the commonalities between similarly spoken and written words. When tested in kindergarten, these children turn out to be the ones with strengths in the phonological domain. For the large group of children with phonological deficiencies who do not understand that the spoken word has segments, and who have not discovered on their own that there is a correspondence between those segments and the segments of the printed word, the current vogue for the so-called (and from our point of view, misnamed) psycholinguistic guessing game and its offshoots, the "whole language" and "language experience" approaches, are likely to be disastrous. Many children taught this way are likely to join the ranks of the millions of functional illiterates in our country who stumble along, guessing at the printed message from their inadequate store of memorized words, unable to decipher a new word they have never seen before.

For those beginners who do not discover the alphabetic principle unaided, an introductory method that provides them with direct instruction in what they need to know is critical (Liberman, 1985; Liberman & Shankweiler, 1979). Direct instruction could begin with language analysis activities that are incorporated into the daily reading lesson. These activities can take many different forms, limited in number and variety only by the creativity of the teacher. The Auditory

Discrimination in Depth Program of Lindamood and Lindamood (1975) is an ingenious method for helping the student to apprehend the internal phonological structure of words It does this by calling the student's attention to the perceived distinctiveness of the articulatory gestures for the various phonemic constituents of spoken words and then demonstrating their sequences in syllables with variously colored blocks. The method was originally developed for individual reading remediation, but is currently being adapted for classroom use.

Adaptations of three exercises that we advocated some years ago (Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980) have recently been shown by one of our colleagues (Blachman, 1987) to be effective in improving reading skills even in an inner city school with a high incidence of reading failure. They are outlined in Figure 1. In the first procedure, one originally devised by the Soviet educator, Elkonin (1973), Blachman presents the child with a simple line drawing representing the word to be analyzed. A rectangle under the drawing is divided into squares equal in number to the phonemes in the picture word. The children are taught to say the word slowly, placing a counter in the appropriate squares of the diagram as the word is being slowly articulated. The words selected must begin with a fricative, liquid, or nasal rather than a stop consonant in order to permit their component phonemes to be accessed readily. Later, as the child progresses, the counters are color-coded—one color for vowels, another for consonants.

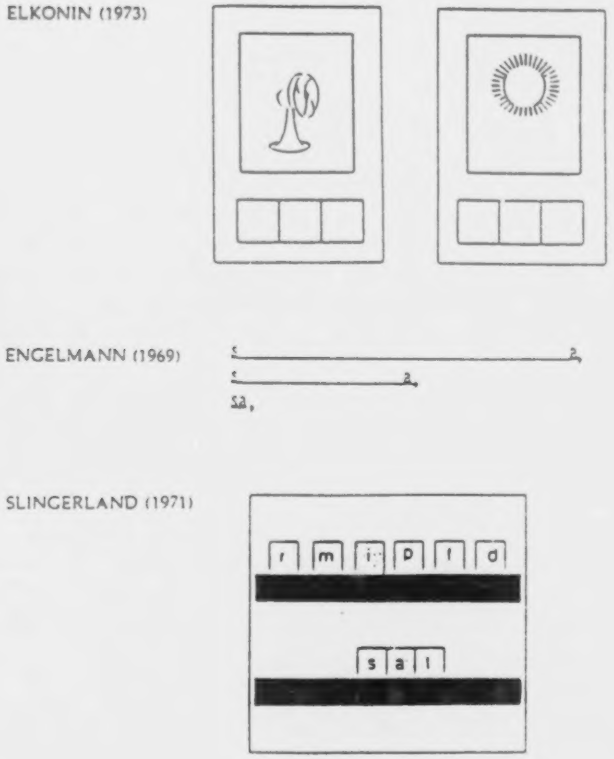


Figure 1. Language analysis activities (after Blachman, 1987 and Liberman et al., 1980).

Letter symbols can be added as well. In another activity, this one adapted from Engelmann (1969), the children are taught how to read as a single unit the combination of a consonant followed by a vowel. For example, the teacher writes a consonant on the blackboard (preferably a fricative, nasal, or liquid)—the letter *s*, for example—and pronounces it, holding it over time until she writes the vowel and pronounces that. The length of time between the pronunciations of the initial consonant and the vowel (as well as a line drawn between them on the board) is then reduced step by step until the two phonemes are pronounced as a single unit—"sa." By adding stop consonants in the final position and pronouncing the resultant words, the children can begin to accumulate a pool of real words (sag, sat, sad, etc.). Thereafter, new vowels and new consonants can be introduced in the same way, and built into new words which are incorporated into stories to be read and written.

A similar effect can be produced by a third procedure, adapted from Slingerland (1971), in which a small pocket chart is used by the child at each desk to manipulate individual letters to form new words and learn new phonemes. The words thus constructed, along with a few nonphonetic "sight" words, can be used in stories and poems to be read and written by the child. Note that the child is now reading and writing words the structure of which is no longer a mystery and the understanding of which can be used productively to form related words (bag, bat, bad, big, bit, bid, etc.).

All these language analysis activities and others like them can be played as games in which the introduction of each new element not only informs but delights. Beginning readers with adequate phonological ability will require only a relatively brief exposure to such activities. They will soon develop skills that will enable them to decode the new words of the text and to go from them to the meaning of the passage. For such readers, language analysis can be quickly followed, or even accompanied, by practice with interesting reading materials from other sources. These children will benefit from the added skill that comes from increased reading practice and the further enhancement of vocabulary and knowledge that comes with expanded reading and life experience. But unless they receive extra assistance, the many beginners with weakness in phonological skills, who may include as many as 20 to 25 percent of the children, will remain locked into a sight-word stage of reading, able to cope only with those few words they have already memorized. They will not learn to decode new words—the essence of true reading skill—unless the method initially includes more intensive, direct, and systematic training in phonological structure and demonstrates how it relates to the way words are written. Research support for this view has been available for at least twenty years (see Chall, 1967 or Pflaum, Walberg, Karegianes, & Rasher, 1980). It is surely time to put the research into practice.

REFERENCES

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory and Cognition, 10*, 451-456.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and activation* (Vol. 4). New York: Academic Press.

Benton, A. L., & Pearl, D. (1978). *Dyslexia: An appraisal of current knowledge*. New York: Oxford University Press.

Blachman, B. (1984). The relationships of rapid naming ability and language analysis skills to kindergarten and first grade reading achievement. *Journal of Educational Psychology, 76*, 610-622.

Blachman, B. (1987). An alternative classroom reading program for learning disabled and other low-achieving children. In W. Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education*. Baltimore: The Orton Dyslexia Society.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature, 301*, 419-421.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics*. New York: Academic Press.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2*, 201-212.

Byrne, B., & Ledez, J. (1983). Phonological awareness in reading-disabled adults. *Australian Journal of Psychology, 35*, 185-197.

Catts, H. W. (1986). Speech production/phonological deficits in reading disordered children. *Journal of Learning Disabilities, 19(8)*, 504-508.

Chall, J. (1967). *Learning to read: The great debate*. New York: McGraw Hill.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies in information processing*. London: Academic Press.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75-84.

Conrad, R. (1979). *The deaf child*. London: Harper & Row.

Content, A., Morais, J., Alegria, J., & Bertelson, P. (1982). Accelerating the development of phonetic segmentation skills in kindergartners. *Cahiers de Psychologie Cognitive, 2*, 259-269.

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G., & Katz, L. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1-16.

Crain, S. (1989). Why poor readers misunderstand spoken sentences. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 133-165). IARLD Research Monograph Series. Ann Arbor: University of Michigan Press.

Crain, S., Shankweiler, D., Macaruso, P., & Bar-Sholom, E. (1990). Working memory and sentence comprehension: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *The neuropsychology of memory disorder*. Cambridge, England: Cambridge University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

Doehring, D. G. (1968). *Patterns of impairment in specific reading disability*. Bloomington: Indiana University Press.

Elkonin, D. B. (1973). U. S. S. R. In J. Downing (Ed.), *Comparative reading*. New York: MacMillan.

Engelmann, S. (1969). *Preventing failure in the primary grades.* Chicago: Science Research Associates.

Fox, B., & Routh, D. K. (1980). Phonetic analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9,* 115-119.

Gleitman, L. R., & Rozin, P. (1977). The structure and acquisition of reading. Relations between orthographies and structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* Hillsdale, NJ: Erlbaum.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Experimental Psychology, 68,* 680-688.

Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association.

Hanson, V. L. (1989). Phonology and reading: Evidence from profoundly deaf readers. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 69-89). IARLD Research Monograph Series. Ann Arbor: University of Michigan Press.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 572-583.

Hanson, V. L., & Fowler, C. A. (1987). Phonological coding in word reading: Evidence from hearing and deaf readers. *Memory and Cognition, 15(3),* 199-207.

Hanson, V. L., Liberman, I. Y., & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. *Journal of Experimental Child Psychology, 37*, 398-393.

Helfgott, J. (1976). Phoneme segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. *Contemporary Educational Psychology, 1,* 157-169.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition, 22,* 25-257.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32,* 474-484.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press

Liberman, A. M. (1989). Reading is hard just because listening is easy. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Wenner-Gren International Symposium Series 54: Brain and Reading*. Hampshire, England: Macmillan.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Liberman, A. M., & Mattingly, I. G. (1989) A specialization for speech perception. *Science, 243,* 489-494.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Winston.

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society, 21,* 71-87.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23,* 65-77.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), *Progress in learning disabilities (Vol. 5).* New York: Grune & Stratton.

Liberman, I. Y. (1985). Should so-called modality preferences determine the nature of instruction for children with learning disabilities? In F. H. Duffy & N. Geschwind (Eds.), *Dyslexia: A neuroscientific approach to clinical evaluation*. Boston: Little, Brown.

Liberman, I. Y., Mann, V., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18,* 367-375.

Liberman, I. Y., Rubin, H., Duques, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. In J. Kavanagh & D. Gray (Eds.), *Biobehavioral Measures of Dyslexia*. Parkton, MD: York Press.

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. B. Resnik & P. A. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, NJ: Erlbaum.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6,* 8-17.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. (1980). Steps toward literacy. In P. Levinson & C. Sloan (Eds.), *Auditory processing and language: Clinical and research perspectives*. New York: Grune and Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Lindamood, C. H., & Lindamood, P. C. (1975). *DLM Teaching Resources*. Allen, Texas.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21,* 159-173.

Lundberg, I., Frost, J., & Petersen, O-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23/3,* 263-284.

Marcel, A. (1980). Phonological awareness and phonological representation: Investigation of a specific spelling problem. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Magnusson, E., & Naucler, K. (1987). Language disordered and normally speaking children's development of spoken and written language: Preliminary results from a longitudinal study. *Reports from Uppsala University, Linguistics Department, 16,* 35-63.

Mann, V., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory and Cognition, 8,* 329-335.

Mann, V., Shankweiler, D., & Smith, S. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11,* 627-643.

deManrique, A. M. B., & Gramigna, S. (1984). La segmentacion fonologica y silabica en ninos de preescolar y primer grado. *Lectura y Vida, 5,* 4-13.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech arise spontaneously? *Cognition, 7,* 323-331.

Morais, J., Cluytens, M., & Alegria, J. (1984). Segmentation abilities of dyslexics and normal readers. *Perceptual and Motor Skills, 58,* 221-222.

Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology, 24,* 35-44.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior, 14,* 33-42.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading, Volume 1*.

Pflaum, S. W., Walberg, H. J., Karegianes, M. L., & Rasher, S. P. (1980). Reading instruction: A quantitative analysis. *Educational Research, 9,* 12-18.

Pratt, A. (1985). *The relationship of linguistic awareness to reading skill in children and adults*. Unpublished doctoral dissertation, University of Rhode Island.

Read, C. (1986). *Children's creative spelling*. London: Routledge & Kegan Paul.

Read, C., & Ruyter, L. (1985). Reading and spelling skills in adults of low literacy. *Reading and Special Education, 6,* 43-52.

Savin, H. (1972). What the child knows about speech when he starts to learn to read. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Shankweiler, D. (1989). How problems of comprehension are related to difficulties in decoding. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 35-68). IARLD Research Monograph Series. Ann Arbor: University of Michigan Press.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorders: a modular approach. *Cognition, 24,* 139-168.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 531-545.

Shankweiler, D., Smith, S. T., & Mann, V. (1984). Repetition and comprehension of spoken sentences by reading disabled children. *Brain and Language, 12,* 241-257.

Slingerland, B. H. (1971). *A Multisensory Approach to Language Arts for Specific Language Disability Children: A Guide for Primary Teachers*. Cambridge, MA: Educators Publishing Service.

Smith, F. (1971). *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York: Holt, Rinehart, & Winston.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S. (1989). Syntactic comprehension in young poor readers. *Applied Psycholinguistics, 10,* 429-454.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: I. Word coding. *Journal of Learning Disabilities, 15,* 449-572.

Stein, C. L., Cairns, H. S., & Zurif, E. B. (1984). Sentence comprehension limitations related to syntactic deficits in reading disabled children. *Applied Psycholinguistics, 5,* 305-322.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuo-motor structure. In A. Allport, D. MacKay, W. Prinz, & E. Sheerer (Eds.), *Language perception and production*. London: Academic Press.

Treiman, R. A., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Walker (Eds.), *Reading research: Advances in theory and practice, 3*. New York: Academic Press.

Vellutino, F. R. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.

Vellutino, F. R., & Scanlon, D. (1987). Phonological coding and phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33/3,* 321-363.

Vernon, M. D. (1957). *Backwardness in reading, a study of its nature and origin*. Cambridge, England: Cambridge University Press.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Waters, G. S., Seidenberg, M. S., & Bruck, M. (1984). Children's and adult's use of spelling-sound information in three reading tasks. *Memory & Cognition, 12,* 293-305.

FOOTNOTES

[*] In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and Reading Disability: Solving the Reading Puzzle*. IARLD Research Monograph Series. Ann Arbor: University of Michigan Press. (1989).

[†] Also University of Connecticut, Storrs.

[††] Also Yale University and University of Connecticut.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 14-27

Language Development from an Evolutionary Perspective*

Michael Studdert-Kennedy†

EVOLUTION AND DEVELOPMENT

Preliminary

In his famous course of lectures at the University of Geneva (1906-1911), de Saussure distinguished *langue*, language as a system, a cultural institution, from *parole*, language as spoken and heard by individuals: "...language is not complete in any speaker; it exists only within a collectivity...only by virtue of a sort of contract signed by members of a community" (de Saussure, 1966, p. 14). Language was thus seen as an abstract property of a group, related to its variable individual speakers somewhat as a species is to its variable individual members.

Nineteenth century attempts to apply evolutionary principles to language did indeed view each language as a species. For example, Darwin wrote: "The formation of different languages and of distinct species...[is] curiously parallel...We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation" (1871, pp. 465-466). Here, Darwin was following in the steps of August Schleicher (1821-1868), who drew on the earlier work of Darwin himself and on the vast scholarship of 19th century European philology, to construct a taxonomy of Indo-European languages. Later work has continued to draw on Darwinian principles to construct evolutionary trees of language families (Lehmann, 1973).

However, a strict analogy between languages and species is untenable for both linguistic and biological reasons. For the historical linguist the most obvious difficulty is that a biological species is a reproductively isolated population: properties of one species cannot pass to another. Yet languages in contact clearly do influence one another, even after they have become discrete sociopolitical entities. Often we cannot then know how far shared properties have arisen from contact, how far from common descent.

A more serious difficulty follows from two central tenets of neo-Darwinian theory (the modern synthesis of Mendelian genetics with Darwinian natural selection). First, in the evolution of a species, the principal unit of variation and selection is not the species, but the individual organism. Individuals within a species differ in the number of offspring they produce; genes whose expression increases the relative number of an individual's offspring will increase their own relative number in the species' gene pool. The characteristics of a species are therefore determined by competition among individuals of that species. And, by analogy, the characteristics of a language are formed by competition among speakers of that language. So far, so good.

The difficulty arises when we combine the principle of individual selection with the so-called central dogma of modern biology, the "Weismann barrier," insulating germ cells from body cells: genes alter body cells, but body cells cannot alter genes. In other words, biological evolution does not proceed by the transmission of acquired characters across generations, and this is precisely what an evolutionary model of language change requires. We must therefore distinguish the cultural, or Lamarckian, evolution of language, a concern of historical linguistics, from its biological, or neo-Darwinian, evolution, a concern of developmental biology.

This distinction was first clearly formulated by Chomsky (1965, 1986), who recognized that de Saussure's definition of language as a property or product of a social group did not lend itself to treatment in biological evolutionary terms. Chomsky made the necessary move by reformulating the *langue-parole* distinction as *competence* (what a speaker-listener knows) and *performance* (what a speaker-listener does in implementing knowledge). He thus set the locus of human language capacity in the individual's mind/brain. He also set the stage for the modern study of language acquisition.

Form and function

Nonetheless, the competence-performance distinction as usually formulated also does not lend itself to evolutionary treatment. Competence is knowledge of a particular language, formed by interaction between an innate schema, the "universal grammar," and the grammar of the language being learned. Universal grammar is "...a theory of the 'initial state' of the language faculty, prior to any linguistic experience" (Chomsky, 1986, pp. 3-4). Performance is thus said to be a product of a partially innate competence. In what follows, I shall argue that this precisely reverses the true course of development. Universal grammar is not a prescription, or program, for development, but a partial and *a posteriori* description of the phenotypic product of the developmental system (cf. Oyama, 1985). Universal grammar is a consequence, not a condition of development.

What is at stake here is the relation between form (competence) and function (performance). Zoologists have traditionally stressed the harmonious match between form and function, as expressed in an animal's mode of life, but they have disagreed on how the match comes about. From classical times down into the 19th century, the standard belief was that species and genera had fixed, unchanging forms or essences: departures from the species prototype were "unreal," and structure took precedence over function. For example, according to Mayr (1982), Georges Cuvier (1769-1832), the great French zoologist, held that "...structure has primacy over function and habit, and ...only a change in structure might necessitate a change in function" (p. 367). Certain non-Darwinian French zoologists still hold such views: "...evolution originates in parent forms; if these are absent, new types of organization never appear" (Grassé, 1977, p. 75). Thus, Chomsky and his structuralist forebears in linguistics align themselves with the essentialist tradition in biology by asserting the primacy of form over function.

For Darwin and modern evolutionary biologists, working within the British empiricist tradition, the form-function relation is reversed. Species are not eternally fixed. A species is a genetically variable population of individuals, adapted to a particular ecological niche and thereby reproductively isolated. No aspect of an animal's structure determines a unique function. Rather, a structure determines an unbounded range of functions, to some of which it is more nicely tailored than others. (We can use a screwdriver to drive screws, or as a dagger, a lever, a drumstick, a fork...). If certain members of a species are forced by competition with their fellows, or by an environmental change, into a new mode of life, a new habitat, they may call on some hitherto unused potential function of their structure. The new mode of life then confers a reproductive advantage on those individuals whose structure is marginally better suited (due to small differences in their genetic make-up) to the new mode. Little by little, as with the beaks of Darwin's famous finches (Lack, 1947), the new selection pressures reshape the old structure to its new function. Behavior is thus the great "pacemaker of evolutionary change" (Mayr, 1982, p. 612), the cause not the consequence of speciation and of species form.

In short, a commitment to gradual evolution by natural selection entails a commitment to the primacy of function over form. My central assumption, then, is that language competence (the neural substrate of language form) was shaped in phylogeny, and is still shaped in ontogeny, by language performance (function, behavior).

Behavior as the pacemaker of development

If we extend Mayr's dictum on evolution to development, we see development in a new light. We are freed from preoccupation with the "initial state" as an index of genetic endowment, and from the nativist-empiricist controversy that has dominated studies of language development almost since their inception.

At birth, an organism suffers an abrupt change in the quality and quantity of environmental conditions that may affect its growth. But the discontinuity does not change the developmental system. Gene action does not cease at birth, nor does regulation of gene action by the cellular environment. Rather, the developmental circle widens. The new external environment affords a broader range of stimulation, eliciting a broader range of response. Changes in the organism's environment and behavior now mediate, to a steadily increasing degree, changes in the cellular environment that controls genetic action.

The developmental course is not one of simple maturation (cf. Borer & Wexler, 1987). Language development will not go forward merely because the environment meets general conditions of survival—air, food, other people, and so on. Specifically linguistic input and, probably, output are called for so that if either is set to zero, development stops. Specific extralinguistic processes of perceptuomotor and cognitive development—a growing grasp on the physical and social modes of being and acting that language represents—must also feed into the language system. Thus, the proper study of language growth is the sequence of behavioral and cognitive conditions, linguistic and extralinguistic, that precipitate language change. The task of this "post-natal embryology," as we might call it, is to chart the course by which perceptual and motoric functions induce structure, from undifferentiated infant performance to differentiated adult competence.

In short, I am proposing that principles of growth, generally accepted in the development of the peripheral anatomy, also apply to the central nervous system. For example, writing of the plasticity of growing bones, D'Arcy Thompson (1917/1961) remarks "...the very important physiological truth that a condition of *strain*, the result of a *stress*, is a direct stimulus to growth itself. This indeed is no less than one of the cardinal facts of theoretical biology. The soles of our boots wear thin, but the soles of our feet grow thick the more we walk upon them: for it would seem that the living cells are 'stimulated' by pressure, or by what we call 'exercise,' to increase and multiply" (p. 238, italics in the original). Bosma (1975) invokes this principle of exercise in describing the development of the vocal apparatus. Perhaps the development of the neural representations of speech and language is also an instance of "...one of the cardinal facts of theoretical biology."

Ontogeny (sometimes) parallels phylogeny

Complex functions, and the physical structures that support them, arise in evolution by gradual differentiation from simpler forms. Every evolutionary change is a change in development that is preserved in later generations. Accordingly, when a complex function is an evolutionarily coherent, hierarchically developed system, ontogeny may parallel, or recapitulate, phylogeny. I should emphasize that I am not proposing to reinstate the discredited "biogenetic law" of Ernst Haeckel (1834-1919). I am merely drawing attention to developmental facts well-known since the embryological studies of Karl Ernst von Baer (1792-1876) in the early 19th century (Gould, 1977).

Of course, if we lack, as we largely do for language, precursor forms that confirm the sequence, we cannot be sure that the order of development we now observe was the actual order of evolution. Such precursors as we may establish—for example, lateralized systems for neural control of body posture and manual function (MacNeilage, in press), or certain capacities for symbolic representation (Savage-Rumbaugh, in press)—do not help us in the present context. Some "stages" may have been inserted into the evolutionary sequence later than others that now follow them in development. This may be true, for example, of left hemisphere sensitivity to speech at, or soon after, birth because it is unlikely that a specialized neural substrate for speech evolved before speech itself.

Yet other processes may have evolved independently, parallel to processes with which they were later integrated. For example, some form of segmental phonology, affording at least a modest lexicon, would seem necessarily to have evolved before syntax began to take shape, and we still observe this sequence in development. But later stages of phonology and syntax may have evolved, as they still seem to develop, more or less independently. Similarly, prosodic variations in pitch, amplitude and duration, characteristic of both human and non-human systems of communication, perhaps first followed an independent course of evolution, to be modified and integrated into the linguistic system only as longer utterances and more finely differentiated syntactic functions emerged. The double dissociation of right and left hemispheres for emotional and linguistic uses of prosody may reflect such a course of evolution.

Yet further limits on a phylogenetic interpretation of language ontogeny may seem to arise because the child is born into a community of companions who already speak a language. Moreover, the child is born with a vocal tract and a pair of hands that soon mature into forms adaptable for spoken or signed language, and with a rich, plastic neural substrate fit for shaping by cognitive and linguistic function. The exogenous and endogenous conditions of development therefore differ radically from those that must have prevailed in a hominid community where language was not yet fully formed.

Nonetheless, language acquisition is not instantaneous. The child does not have immediate access to the full adult language that surrounds it.

The effective linguistic environment changes, step by step, as the child comes into possession of new cognitive and linguistic capacities. The problem of how linguistic input is ordered so as to ensure coherent development is solved by the child's own increasingly differentiated linguistic attention. With each new step the child finds the next step waiting, as it were, in the adult language because the adult language is adapted to the child no less than the child to the language. The reason for this is simply that language, like every evolved form, is the product of successive ontogenies, its structure a record of its own evolution (cf. Locke, 1983). Looked at in this way, language is not an object, or even a skill, that lies outside the child and has somehow to be acquired or internalized. Rather, it is a mode of action into which the child grows because the mode is implicit in the human developmental system.

We conclude that language, as a complex, hierarchical, behavioral structure with a lengthy course of development, is a good candidate for (circumspect) study in a recapitulatory framework because its development is rich in sequential dependencies: syllables and formulaic phrases before phonemes and features (as I shall argue below), holophrases before words, words before simple sentences, simple sentences before lexical categories, lexical categories before complex sentences, and so on. Thus, if we assume that each of the subsystems, phonology and syntax, evolved hierarchically by repeated cycles of differentiation and integration, we may recover their course of evolution by tracing the course of their growth. The general heuristic value of the assumption is that it not only throws light on evolution, but also promises an understanding of development in functional terms.

Development is not teleological

If we combine the principles of gradual evolution and of a limited, functional recapitulation, we are freed from the temptation to assign purpose to development. At each point in its development an organism is already complete, adapted and adapting, as best it can, to present conditions, internal and external. Just as earlier evolutionary forms existed for themselves, not for any later forms to which they might give rise, so the present form of a developing organism has its own present function. A child does not learn its first words so that it may later combine them into sentences. First words have their own economy. More generally, a child's grammar, at each stage of development, is a possible adult grammar.

This last statement is limited exactly to the extent that the principle of recapitulation is limited. Some stages may have been inserted into the developmental sequence relatively late in evolution. Others may now serve a new function, having lost the function for which they originally arose. Such stages, perhaps particularly during the early months of rapid growth before linguistic functions have begun to differentiate, would have survived through successive generations because they facilitated later adaptive changes, while not infringing on present functions. In this sense, an early form may "preadapt" for a later. However, it would be an error to see such preadaptations as literal preformations. Development, like evolution, is a tinker, putting to present use whatever chance has laid to hand (cf. Jacob, 1977).

Individuals reach the same developmental ends by different routes and at different rates

Many linguists believe that universal aspects of language are "innate." Universals are said to be purely genetically determined, and either to be present at birth (Chomsky, 1986) or to mature without contribution from the environment (Borer & Wexler, 1987). Aspects that vary across languages, though perhaps constrained to a few values by principles of universal grammar, are said to be learned from, or "fixed" by, the surrounding language. This is the model of "parameter setting" in syntactic theory (Chomsky, 1981; Roeper & Williams, 1987) and, in effect, of Jakobson's (1968) account of phonological development (cf. Goad & Ingram, 1987).

I do not challenge the descriptive adequacy of such theories. What I question is the proposed developmental mechanism, the attribution of species invariance to the genes, within-species variability to the environment. For, in fact, both invariance and variability arise from both genes and environment. The point is important because individual similarities in language development are often taken as support for a "biological" account of the process, while individual differences are seen as a threat (Goad & Ingram, 1987).

Consider, first, that identical phenotypes may be induced either by a genetic change in an unchanged environment or by a changed environment acting on an unchanged genome. For example, a yellow (as opposed to grey) body in the fruit fly (*Drosophila melanogaster*) can be produced either by mutation or by feeding the fly larva on fruit jelly impregnated with yellow dye (Dobzhansky, 1957). Similarly, the lack of a posterior cross-vein in fruit fly wings can result either from mutation or from subjecting the fly pupa to a brief high-temperature shock early in development (Waddington, 1975, Chapters 7 and 8). Another example, not due to experimental intervention, comes from Piaget (1978) who studied fresh water snails in Swiss lakes during the late 1920's. He describes three species. One species (*Limnaea stagnalis*), found in deep, calm waters, has an elongated shell; two other species (*L. lacustris* and *L. bodamica*), found in shallow, rough waters, have almost indistinguishable stubby, contracted shells. If the two stubby species are bred in the calm water of a laboratory aquarium, *L. bodamica* retains its stubby shell over many generations, while *L. lacustris* gradually takes on the elongated form characteristic of *L. stagnalis*. Had Piaget simply described the two stubby species, he might have attributed their similarity to shared genes. But we cannot reliably infer genotype from phenotype, as discussions of language development often assume. Without controlled breeding studies, we cannot determine whether an invariant, species-typical form reflects genetic constancy or environmental constancy.

Notice, moreover, that even when we have isolated genetic or environmental factors that contribute to some aspect of phenotypic variation, we have not demonstrated that the variable held constant has no effect. The snail species that takes an elongated form after generations in calm water, a contracted form after generations in rough water, can adapt to the environmental change because it is genetically equipped to do so. The snail species that retains its contracted form, whether bred in rough waters or in calm, lacks this genetic potential. But we cannot conclude from this that the environment has no effect on the form of its shell. We can conclude only that the two environments to which the snail was exposed did not suffice to select for a change in form although other environments might have done so.

In short, phenotypic form is both genetically and environmentally determined, a fact of some importance to language development. If we cannot assign phenotypic invariance, even in these "simple" examples, either to genes alone or to environment alone, we are surely not justified in doing so for language universals, or for the within-language constancies in linguistic development that are often cited as evidence for an "innate" language competence.

By the same token, variability in development among children learning the same language cannot be assigned solely to the environment. Certainly, different children, even within the same social class, must be exposed to widely varying patterns and frequencies of linguistic input, and such differences are likely to affect the course and rate of development. At the same time, children differ genetically, and these differences too affect development (Locke & Mather, in press). The problem is not to assign variability to a genetic or environmental source, but to understand how children resist the effects of genetic and environmental variation so as to arrive at a common language.

Development is buffered against extreme variation

The problem of uniformity within and across languages is, in some respects, the obverse of the problem of language diversity. Languages, like species, differ because individuals differ. If all members of a species were genetically identical, followed an identical course of development, and arrived at an identical developmental term, new species could never arise. Genetic variation within a species, affording subtle individual differences in adaptive response to new environments, is the basis for the origin of species by natural selection. Similarly, if all speakers of a language followed an identical course of development to an identical linguistic term, new languages could never arise.

Of course, differences among speakers that lead to language change are unlikely to rest on genetic differences, because speakers of different languages do not differ systematically in their language-related genetic endowment: children learn any first language to which they are exposed. In fact, from a genetic point of view, language differences rest on commonalties rather than differences: the shared capacity to adapt to any linguistic environment and to learn the surrounding language. In other words, it is precisely because language is supported by (we must presume) many thousands of genes which depend on the environment to trigger their expression, that languages can be learned—and that languages differ. Subtle, environmentally induced individual differences in language development, selected and transmitted to other individuals by sociocultural forces of which we understand at present, very little, are one source of language diversity.

What then are the forces that resist change, guiding speakers into a common language? The problem is a special case, analogous to the general problem in developmental biology of how species maintain their identities across generations. We can gain conceptual leverage on the question by invoking the principle of "canalization" (Waddington, 1957, 1975). The term is purely descriptive, intended to capture "...a large number of well-known facts in genetics and embryology, all of which are summarized in the statement that the development of any particular phenotypic character is to some extent modifiable, and to some extent resistant to modification, by changes either in the genotype or in the environment" (Waddington, 1975, p. 72). In other words, both genetic and environmental variations are buffered against extreme divergence, canalized toward equifinality by constraints from the developing system. Genes and environment act in reciprocal, mutually constraining concert to assure a stable trajectory of growth.

To characterize the constraints on language growth in any detail is well beyond the scope of present knowledge. There would seem, however, to be two main sources of constraint, one in the child itself, one in the physical, social and linguistic environment. Within the child, each step in cognitive and linguistic development opens new paths and closes others, an automatic consequence of the selection, and increasing differentiation, of the neural structures that support language and cognition, continuing the processes of tissue differentiation in the embryo. Once launched, the learning of a particular language becomes increasingly constrained by structural changes in the child's brain. These changes open paths into the language being learned, and perhaps into other languages of the same general type. Presumably, development along these paths will cease if the child is abruptly transferred to a markedly different linguistic environment. (Studies of "savings," that is, of the head-start or lack of it, displayed by children who have switched languages or been exposed to a second language, might then throw light on cognitive and linguistic commonalties between the languages and on their inferred neural substrates.)

The external constraints on the child arise, first, from a general social context that invites the child to engage with its companions and to match its behavior to theirs. The second source of constraint is the language that the child's companions speak. Every language is a solution, one of an uncountable, but presumably limited, set of solutions to the problem of developing a communicative system within the perceptuomotor, memorial and cognitive limits of humans (cf. Lindblom, 1983). The child is then easily guided into its language because, as we have already remarked, every language has evolved under constraints very like those that limit the child.

Given the increasing strength of endogenous and exogenous constraints as the child grows, we would expect phenotypic variation in phonology and syntax, if not in lexicon, to diminish as children come into possession of their language. Nonetheless, individual variability, far from being evidence against a biological account of language development, is a hallmark of developing biological systems.

In the following section, I will attempt to illustrate some of the principles sketched above as they might apply to the development of phonological form over the first two years of life.

THE EMERGENCE OF PHONETIC SEGMENTS

Perception

The first systematic studies of infant capacity to perceive speech derived from the well-known studies of adult categorical perception. In a seminal set of experiments, the model for many others, Eimas and his colleagues (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) demonstrated that 1- and 4-month old infants displayed the same pattern of discrimination between syllables on a synthetic voice onset time (VOT) continuum as adults: they discriminated significantly better between syllables belonging to different English phoneme categories than between syllables belonging to the same phoneme category. Much the same result came from later infant studies of stop consonant place of articulation, consonant manner, the [r]-[l] distinction, and almost every other synthetic continuum on which infants were tested (see Aslin, Pisoni, & Jusczyk, 1983, for a comprehensive review).

From this accumulated wealth of evidence Eimas (1975) concluded that categorical perception reflects the operation of innate mechanisms, specialized for processing speech, and that "...these early categories serve as the basis for future phonetic categories" (p. 342). This conclusion is often cited by students of language acquisition whose primary interest is in syntax. For example, Gleitman and Wanner (1982) summarily dismiss the problem of the origin of discrete phonetic segments (a problem no less severe than the problem of the origin of morphemic units to which they devote much attention) by citing the work of Eimas and others to support the claim that "...no learning apparatus is required for an initial segmentation of the acoustic wave into discrete phones. The segmentation has been provided in the nervous system" (p. 16). Exactly how this capacity for segmentation came to be evolutionarily, or comes to be ontogenetically, "provided in the nervous system," they do not consider.

In any event, many reasons to doubt Eimas's interpretation of the infant data have been given by Jusczyk (1981), Kuhl (1987), Studdert-Kennedy (1986), and Walley, Pisoni and Aslin (1981). I will not rehearse the reasons in detail, but here briefly note three. First, we now know that categorical perception is peculiar to neither speech nor audition (see Harnad (1985) for several relevant papers). Second, Kuhl and her colleagues have demonstrated categorical effects on labial, alveolar and velar VOT continua in chinchillas (Kuhl & Miller, 1975), and on a place continuum in macaques (Kuhl & Padden, 1982, 1983). Thus, whatever the bearing of categorical studies on the development of speech perception, they clearly do not reflect a perceptual specialization. Third, the claim that categorical perception reveals the basis for future phonetic categories confuses two types of category. The categories mimicked by a synthetic series vary along a single acoustic dimension in a fixed context; they comprise, at most, the random variations that we might observe in a single syllable, spoken repeatedly with identical stress and at an identical rate by the same speaker. However, the phonological categories that a child must form are equivalence classes of intrinsic and extrinsic allophonic variants, formed by execution of a particular phoneme in a range of phonetic contexts, spoken with varying degrees of stress, at different rates, and by different speakers.

For study of the infant's potential grasp on these equivalence classes, we must turn to another body of research. In a systematic series of studies Kuhl and her colleagues have demonstrated that 3- to 6-month old infants can learn to recognize the equivalence of: (i) isolated vowels spoken by a male, a female or a child; (ii) syllables with the same consonantal onset before different vowels; (iii) syllables with a particular consonantal acoustic pattern in initial, medial or final position; (iv) syllables that share an initial "feature" (stop, nasal). (For a comprehensive review, see Kuhl, 1987). Yet we should be cautious in interpreting even these studies as evidence that infants can recognize phonetic (as opposed to acoustic) invariants, if only because Kluender, Diehl and Killeen (1987) have successfully trained Japanese quail to form equivalence classes across syllables spoken with different pitch contours, by different speakers or with the same consonant before different vowels. In short, we have no grounds for claiming that perceptual analysis of the acoustic structure of a spoken syllable, and the formation of categories from the resulting components, engage distinctively human capacities, whether infant or adult.

In fact, these infant studies seem to have no more than a general bearing on the specialized development of language. They are psychophysical studies, demonstrating that infants at, or soon after, birth have the capacity to discriminate and categorize certain acoustic patterns which occur in speech. They provide detailed support for a general observation, suggested by the fact that speech sounds are concentrated in the few octaves of the acoustic spectrum to which humans (and many other animals) are most sensitive: spoken language has evolved and develops within the constraints of prelinguistic auditory capacity. Surely, it would be surprising if this were not so. We do not expect an animal to have a communication system that is *not* matched to its sensory capacities.

However, the most serious objection to the standard interpretation of these studies is that by assuming the child to be innately endowed with sensitivity to the smallest phonetic units of which spoken utterances are composed, they implicitly adopt a view of development as proceeding from the specific to the general rather than the reverse. Yet, as we have already seen, it is a general rule of both phylogeny and ontogeny that complex structures evolve by differentiation of smaller structures from larger. Accordingly, we should not expect children to build words from phonemes as adults do; rather, we should expect phonemes to emerge from words. We may note, in passing, that a similar principle must apply to the development of word classes and syntactic structures, a fact not generally recognized in developmental psycholinguistics.

From this vantage we can take a different view of infant speech perception than has hitherto prevailed, one that emphasizes *development of function* rather than psychophysical capacity. One aspect of this work will entail charting the process of "attunement" to the surrounding language, the course by which the infant *learns to perceive* speech, bringing to bear capacities evidenced in the laboratory on the speech it hears at home. Work along these lines has, in fact, already begun, with studies of infants' loss of sensitivity to phonetic contrasts not deployed in their native language (e.g., Best, McRoberts, & Sithole, 1988; Werker & Tees, 1984). Other studies by Hirsh-Pasek and her colleagues (e.g., Hirsh-Pasek & Golinkoff, in press; Hirsh-Pasek, Nelson, Jusczyk, Cassidy, Druss, & Kennedy, 1987) are tracing the development of sensitivity to prosodic patterns that specify clausal units, perhaps the thin end of an infant wedge into syntactic structure.

Here, however, I wish to pursue another aspect of perceptual function: its role in guiding the development of production. For I assume that the development of speech depends not only on the maturation and use of the vocal motor system, but also on the infant's gradual discovery of structure in the speech it hears: By learning how to listen the infant learns how to speak. The speech signal therefore comes to specify for the infant the actions—or, more exactly, the motoric components of articulatory action—by which speech is produced. Yet neither phonemes nor features, the perceptual units typically posited in infant studies, can be defined either acoustically or motorically. They are abstract units beyond the reach of an infant who does not yet know a language. If we are to understand speech development, we must couch our descriptions and frame our experiments in terms of auditory and motoric units to which an infant might reasonably be expected to have access. In what follows I will briefly sketch a speculative account of the process by which infants gradually harness prelinguistic motoric elements to communicative use.

Production

We may discern at least four broad stages in the early development of speech, successive cycles of differentiation and integration that carry the infant from prelinguistic cries and mouthings to the emergence of phonetic segments and the beginnings of phonology in early words. The stages form a necessary hierarchical sequence from the general to the specific, from the nonlinguistic to the linguistic. However, they are not sharply delimited: processes that begin in one stage may continue more or less unchanged into several later stages—for example, a child may continue to babble long after it has begun to produce words. Also, we must be cautious in characterizing a stage as differentiative or integrative for two reasons. First, because differentiation at one level (for example, breaking a syllable into its component gestures) entails integration at another (coordinating movements to form gestures), and vice versa. Second, because even within a level both differentiation and integration go on during a single stage: for example, integrating gestural patterns into consonants and vowels entails simultaneously differentiating a syllable into its segmental components.

Nonetheless, since even a coarse and tentative taxonomy of development may lend insight into its process, let us consider the following four stages: (1) early vocalizing: differentiation of respiratory and vocal tract activities into patterns of soundmaking associated with different nonspeech actions (0-7 months); (2) canonical babbling: integration of nonspeech movement patterns into rhythmic syllabic structures (7-10 months); (3) variegated babbling and early words: differentiation of the syllable into its component gestures (10-15 months); (4) integration of recurrent gestural patterns into canonical phonetic segments (15-24 months). The suggested time periods are, of course, approximate, because children differ widely in the time courses of their development.

Differentiation: Early vocalizing

Over the first half year of life infant sounds progress systematically from clearly nonspeech vocalizations to canonical babbling that invites phonetic transcription. The changing forms are presumably determined, at least in part, by general maturation and by exercise of the vocal apparatus and of its neural control structures (for review, see Kent, 1981). Stark (1986) divides the development of soundmaking over the first 7-8 months of life into three stages: reflexive crying and vegetative sounds (0-8 weeks), cooing and laughter (8 to 20 weeks), and vocal play (16-30 weeks). These stages are of interest in the present context because they reflect changes in the topography of vocal tract activities and in the forms of the resulting sounds by which prelinguistic elements are differentiated and marshalled for protolinguistic use (cf. Oller, 1986).

Reflexive crying, like the distress calls of other animals, has, of course, a communicative (though non-linguistic) function. Cries, executed with a relatively unconstricted vocal tract, are predominantly voiced, often with the formant structure of low to mid front vowels. Vegetative sounds—the grunts, sighs, clicks, stops and pops, associated with breathing and feeding—may be either voiced or voiceless, formed with either an open, vowel-like or a constricted, consonant-like configuration of the vocal tract. The cooing, or comfort, sounds of Stark's second stage tend to occur in a series of 3-10 segments, each of about 500 ms, separated by voiceless intakes of breath and glottal stops. Their energy is concentrated in frequencies below about 1500 Hz, a pattern that Oller (1980) has termed a "quasi-resonant" nucleus, having the form of a nasal consonant or nasalized vowel.

Thus by the end of the fifth month, the infant's sound repertoire already contains a variety protoconsonantal and protovocalic elements. The important changes in the next stage are not so much in the size and quality of the repertoire as in its function and organization. First, sounds begin to lose their original functional moorings: they become "...divorced from their previous cry, vegetative or comfort sound contexts, and are used in a variety of communicative situations" (Stark, 1986, p. 159). Second, sounds become longer (700-1500 ms), and form longer, more complex sequences. The infant emits repetitious strings of consonant-like clicks, trills, friction noises, syllabic nasals with a constriction at the front of the mouth, and lip smackings. Vowel-like sounds, now with a "fully resonant" nucleus (Oller, 1980) often carrying extreme pitch glides, are executed with increased variation in tongue height and front-back placement.

However, the key change during this stage is combinatorial: the infant begins to superimpose movements of tongue, jaw and lips on the laryngeal actions associated with cry (Koopmans-Van Beinum & Van der Stelt, 1986), so that the proportion of supraglottal to glottal articulations gradually increases (Holmgren, Lindblom, Aurelius, Jalling, & Zetterström, 1986.) Even at this early stage, children may differ quite sharply in their preferred types of vocalization (Stark, 1986). Toward the end of the stage increasingly long and complex combinations of tract constrictions and openings appear, forming sequences that Oller (1980) terms marginal babble.

The functional value of the differentiation of early soundmaking into these diverse patterns is not obvious. Teleologically, of course, exercise of the vocal apparatus must contribute to its neural and anatomical development, laying the basis for later integration of sounds into syllables. However, the immediate function may simply lie in the increased range of emotional expression that it affords, with a consequent tightening of dyadic social bonds (cf. Stern, 1985; Trevarthen & Marwick, 1986). Interestingly, MacNeilage (personal communication) has noted a possible precursor of such differentiation in the repetitive girneys (lip smacks, accompanied by a low murmur) exhibited by Japanese macaques in intimate, affiliative situations (Green, 1975).

In any event, I have dwelt on this early stage because it is here that we can see most clearly the tinkering together of disparate non-linguistic patterns of vocal tract activity into the beginnings of phonological structure (cf. Bates et al., in press).

Integration: Canonical babble

With the onset of canonical babble, often a sudden event over a few days in the 7th or 8th month, the infant begins to integrate patterns of vocal tract constriction and opening into unitary, cohesive syllables. Rhythmic, reduplicated sound sequences are common: [bababa], [nenene], [dɪdɪdɪ]. However, the convenient use of phonetic transcription should not mislead us into supposing that the infant has independent control over segments within a syllable. In fact, rhythmic lowering and raising of the jaw in canonical babble seems to occur with little or no independent movement of the tongue (Davis & MacNeilage, 1990). This rhythmic jaw oscillation often begins about the same time as rhythmic movements of the legs and arms (Thelen, 1981; in press), and perhaps facilitates the integration of vocal tract activities into cohesive syllabic patterns.

We owe the clear distinction between early vocalization and canonical babble to Oller (1980, 1986). He describes some half dozen acoustic properties of structures that listeners recognize as canonical, adult consonant-vowel (CV) (constricted-to-open) or CVC syllables. Of these properties the most important in the present context (because they are the only ones that appear rarely, if ever, in early vocalizations) are temporal. The canonical syllable has a duration of no less than 100, no more than 500 ms; onset (and, when present, offset) formant transitions display smooth changes in frequency and amplitude with durations between 25 and 120 ms. These acoustic patterns reflect the infant's increasing skill in integrating the closing and opening phases of jaw movements.

Patterns of movement in babble are not random. During the constriction phase of the syllable, complete closure, as in stops, is favored over partial closure, as in fricatives. Points of closure are biased toward the front of the mouth, engaging lip and tongue tip muscles active in sucking. During the open phase, the favored tongue positions are those associated with low front vowels, indicating that the tongue tends to ride up and down on the jaw with little or no active movement of its own (Davis & MacNeilage, 1990). The glottis is typically approximated throughout the syllable, giving the impression of voiced consonants at syllable onset.

These "phonetic" biases have been reported for infants growing in a number of language environments (Locke, 1983). The biases are also present in many adult languages, perhaps reflecting the infant proclivities from which languages have evolved (Locke, 1983; cf. Lindblom, 1989). Here we have a tangle that cross-linguistic data on babble are still too sparse to resolve. How far do the perhaps universal biases of infant babble reflect the naturational state of the infant vocal apparatus, and how far do they reflect the surrounding language? We may ask much the same question concerning individual differences in babbling repertoire within a language: do they reflect differences in the development of the vocal apparatus or differences in the language patterns which the infants happen to have heard? Whatever the answers to these questions, recent work with deaf infants has shown that the emergence of canonical babble is not purely an effect of maturation.

Deaf infants were once said to exhibit the same babbling patterns as hearing infants, at least over the first year of life (Mavilya, 1969; cf. Locke, 1983). However, recent comparisons have demonstrated that canonical babble does not appear on schedule in deaf infants (Oller & Eilers, 1988; Oller, Eilers, Bull, & Carney, 1985; Stoel-Gammon & Otomo, 1986). Whether this is due to the lack of auditory input from the infant's own vocalizations, from those of its adult companions, or both, we do not know. If self-stimulation were the only essential, a purely maturational account could still hold. To the extent that communicative interchange, or the impulse to imitate the actions of conspecifics plays a role, the onset of canonical babble would be determined, at least in part, by experience of a surrounding language. Persuasive evidence for the role of the surrounding language comes from reports that deaf children exposed to sign language begin to "babble" with their fingers at about the same age as hearing infants begin to babble with their mouths (Newport & Meier, 1985; Laura Petitto, personal communication).

Finally, what is the function of canonical babble? "None" has been the answer of some (e.g., Jakobson, 1968; Lenneberg, 1967) who viewed babble as random mouthing, sharply discontinuous from truly linguistic utterance. However, Locke and Pearson (1988) have recently reported a severe (though not irremediable) delay in the development of speech in a tracheostomized child, deprived of the opportunity to babble from 5 to 20 months. This suggests that babble is a necessary step in normal linguistic development. Moreover, several studies have now shown that babble merges smoothly into early words and that the phonetic structure of a child's first words tends to reflect its babbling preferences (Locke, 1983; Oller, Wieman, Doyle, & Ross, 1975; Vihman, Macken, Miller, Simmons, & Miller, 1985; Vihman & Miller, 1988). The immediate function would seem then to be to continue the "imitative" process of aligning the infant's communicative skills more closely with those of its adult companions.

Differentiation: Variegated babble and early words

Toward the end of the first year full syllable reduplication fades. The infant begins to differentiate the closing and opening gestures of successive syllables so that "variegated" sequences (Oller, 1980) appear, giving the impression of variations in consonant and/or vowel (e.g., /nenɪ/, /mənɚ/, /dɛdi/). Also, at about this time the child produces its first recognizable attempts at adult words or phrases (*All gone, What's that*?), usually (in English children) single syllables or disyllables. The two modes of output then proceed concurrently, often for many months, with words gradually coming to predominate. Over this period the child is gradually forging links between its perceptual and productive capacities.

The perceptual ground for this development seems to be laid by the child's growing attention to words in the surrounding language. Data on early comprehension are scarce, but children evidently accumulate sizeable receptive lexicons before attempting their first words. Benedict (1979) has reported a longitudinal study of comprehension and production in eight children from the age of 9 or 10 months to a point where they had achieved a productive lexicon of 50 words (between 15 and 22 months). In every child comprehension comfortably outstripped production. On the average, the children understood more than 60 words by the time they could produce 10, with a range from 30 to 182 words; and on the average, their receptive lexicon over the period of the study was three times their productive lexicon (Benedict, 1979, Table I).

The gap between perception and production demonstrates that a perceptuomotor link is not "innate," as some have proposed (Liberman & Mattingly, 1985), but must develop. Many studies have shown that the phonetic forms of early words are similar to those of concurrent babble (see Vihman, in press, for review). A child selects for imitation words that match "vocal motor schemes" (McCune & Vihman, 1987) already present in its babble and avoids words that do not match (Menn, 1983). Thus, the child initially grows into a lexicon, as it were, by discovering correspondences between adult words and auditory feedback from its own babbled output. Because the structure of auditory feedback must correspond, in some fashion, to the structure of motor controls that produced it, the child's recognition of babble-to-word correspondences presumably facilitates growth of perceptuomotor links, and their gradual extension to new words.

Each word (or formulaic phrase) seems to be a prosodic unit (Macken, 1979), its production planned as a whole (Menn, 1983). Evidence for this comes from gestural interactions. A child often fails to execute two different places or manners of articulation within the same word, thus maintaining some of the reduplicative tendencies of canonical babble: *dog* [gɑg], *lady* [jeiji], *duck* [tʌt]. Closing and opening gestures also interact. For example, Davis and MacNeilage (1990) report an extensive study of a child's concurrent babbling and speech over the period from 14 to 20 months of age. Their data are replete with instances not only of consonant, but of vowel and even consonant-vowel assimilation. The latter is revealed by the child's preference for high front vowels following alveolar closures and for low, front-central vowels following labial closures. At the same time, these authors also report an inverse relation between consonant and vowel reduplication: where the child succeeds in combating assimilation in the open phases of a disyllable, she often fails to do so in the closing phases, and vice versa. This demonstrates an incipient segregation of consonants and vowels into phonetic classes.

The study by Davis and MacNeilage is particularly important because the child deployed an unusually large lexicon, growing from about 25 words at 14 months to over 750 words at 20 months. Evidently, a child may have a substantial lexicon long before it has fully mastered segmental structure. The principal phonological achievement of this period, then, is internal modification of the integrated syllable by differentiation of its gestural components.

Integration: From gestures to phonemes

The final step in the path from mouth sounds to segments is the integration of gestural patterns of syllabic constriction and opening into the coherent perceptuomotor structures we know as consonants and vowels (Studdert-Kennedy, 1987). As is well known, the status of the segment is problematic. For, on the one hand, we can neither specify the invariant articulatory-acoustic properties shared by all instances of a particular consonant or vowel, nor isolate any given segment as a discrete articulatory-acoustic entity within a syllable. On the other hand, across-word metathetic errors in adult speaking ("Spoonerisms") attest to the functional role of segments in the planning and execution of an utterance (Shattuck-Hufnagel, 1983). Moreover, such errors typically entail exchanges between consonants and vowels that occupy corresponding slots in their respective syllables. Since the exchanging elements may be physically quite disparate, it is evident that their exchange is premised on shared function (onset, nucleus, coda) in the formation of a syllable.

There are therefore two aspects to the emergence of segments as elements of word formation in a child's lexicon. First is the grouping of all instances of a particular gesture-sound pattern into a single class presumably on the basis of their perceptuomotor, or phonetic, similarity (e.g., grouping the initial or final patterns of *dad, dog, bed*, etc. into the class /d/). Second is the distributional analysis and grouping of these gesture-sound patterns into higher-order classes (consonants, vowels) on the basis of their syllabic functions. These processes of phonetic category formation are perhaps analogous to those proposed by Maratsos and Chalkley (1980) for the formation of syntactic categories.

Evidence from across-word metathetic errors for the formation of these classes in young children is sparse. The only systematic data known to me come from Jeri Jaeger (personal communication). She reports her daughter's first across-word metathesis as occurring in her 27th month: *ummy takes* for *tummy aches*. This was followed in her 30th month by *fritty pace* for *pretty face, sea tet* for *tea set*, and *Bernie and Ert* for *Ernie and Bert*. Jaeger did not report data on the size of her daughter's lexicon at this time, nor on the complexity of her multi-word utterances. But the collection of errors suggests that both were well advanced.

Two possible selection pressures may precipitate formation of consonant and vowel classes. One pressure is toward economy of storage as the lexicon increases in size. We have seen that a child may accumulate an appreciable lexicon of some 750 words without showing signs of independent segmental control (Davis & MacNeilage, in press). But this lexicon is roughly one hundredth of the size that it will eventually become; and it seems reasonable to suppose that, as the lexicon increases, words should organize themselves on the basis of their shared gestural and sound properties. Recurrent patterns of laryngeal and supralaryngeal gesture would thus form themselves into unitary classes of potential utility for recognition and activation of lexical items (cf. Lindblom, 1988; Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

A second possible selection pressure is toward rapid lexical access in the formation of multi-word utterances. Several authors (e.g., Branigan, 1979; Donahue, 1986) have argued that the form of early multi-word combinations may be constrained by the child's limited ability to organize and execute the required articulatory sequences. One such constraint might be a child's inability to produce two successive words with different initial places of articulation. Donahue (1986), for example, describes two strategies adopted by her son in his first two-word utterances. One strategy was to attempt only those words that conformed to his preexisting rule of labial harmony: *big book, big bird, big ball* were all attempted, but *big dog* and *big cooky* were "adamantly refused" (p.215). The second strategy was to circumvent the consonant harmony rule by adopting vocalic words lacking consonants (e.g., *where* [ejə], *want* [waɛ]) as pivots that could be comfortably combined with many of the words already in his vocabulary. Such findings imply that the integration of gestures into independent phonemic control structures, or articulatory routines (Menn, 1983), may serve to insulate them from articulatory competition with incompatible gestures and so facilitate their rapid, successive activation in multi-word utterances.

CONCLUSION

I have proposed that the study of language development might be fruitfully cast in an evolutionary and recapitulatory framework, as a sort of post-natal embryology. Language universals (like other species characters) are the endpoint of development (and evolution), a consequence, not a condition, of learning a natural language. Language function thus determines language form, and form is viewed as an *a posteriori* description of, rather than an *a priori* prescription for, development. We are thus freed from the habit of viewing universals as innate, language-specific properties as learned. Both sets of characters are the product of a species-specific developmental system, in which genetic and environmental conditions cannot be separated.

I have attempted to illustrate the approach with a sketch of the process by which the universal phonological categories of consonants and vowels might emerge. They are viewed as deriving by successive cycles of differentiation and integration from prior non-linguistic perceptual and motor capacities.

REFERENCES

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. Haith & J. Campos (Eds.), *Infancy and the biology of development. Carmichael's manual of child psychology, Vol. II* (4th ed., pp. 220-284). New York: Wiley.

Bates, E., Thal, D., & Marchman, V. (in press). Symbols and syntax: A Darwinian approach to language development. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6,* 183-200.

Best, C. T., McRoberts, G. W., & Sithole, N. W. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345-360.

Borer, H., & Wexler, K. (1987) The maturation of syntax. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 123-172). Dordrecht, Holland: D. Reidel.

Bosma, J. F. (1975). Anatomic and physiologic development of the speech apparatus. In D. B. Tower (Ed.), *The nervous system (Vol.3): Human communication and its disorders*. New York: Raven.

Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language, 6,* 411-421.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: M.I.T. Press.

Chomsky, N. (1972). *Language and mind* (rev. ed.). New York: Harcourt Brace Jovanovich .

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures.* Dordrecht, Holland: Foris.

Chomsky, N. (1986). *Knowledge of language.* New York: Praeger Special Studies.

Darwin, C. (1871). *The descent of man*. New York: The Modern Library.

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research, 33,* 17-27.

de Saussure, F. (1966). *Course in general linguistics* (Wade Baskin, Trans.). New York: McGraw Hill.

Dobzhansky, T. G. (1957). *Human genetics.* New York: Columbia University Press.

Donahue, M. (1986). Phonological constraints on the emergence of two-word utterances. *Journal of Child Language, 13,* 209-218.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the /r-l/ distinction by young infants. *Perception and Psychophysics, 18,* 341-347.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Gleitman, L. R., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition* (pp. 3-48). New York: Cambridge University Press.

Goad, H, & Ingram, D. (1987). Individual variation and its relevance to a theory of phonological acquisition. *Journal of Child Language, 14,* 419-432.

Green, S. (1975). Variation of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): A field study. In L. A. Rosenblum (Ed.), *Primate behavior* (Vol. 4, pp. 1-102). New York: Academic Press.

Gould, S.J. (1977). *Ontogeny and phylogeny.* Cambridge, MA: Belknap Press.

Grassé, P-P. (1977). *Evolution of living organisms.* New York: Academic Press.

Harnad, S. (Ed.) (1987). *Categorical perception.* New York: Cambridge University Press.

Hirsh-Pasek, K., & Golinkoff, R. M. (in press). Language comprehension: A new look at old themes. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269-286.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterström, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 51-63). New York: Stockton Press.

Jacob, F. (1977). Evolution as tinkering. *Science, 196,*1161-1166.

Jakobson, R. (1968). *Child language, aphasia and phonological universals* (A. R. Keiler, Trans). The Hague: Mouton

Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kent, R. D. (1981). Sensorimotor aspects of speech development. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of Perception Volume 1* (pp. 161-189). New York: Academic Press.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237,* 1195-1197.

Koopmans-van Beinum, F. J., & Van der Stelt, J. M. (1986). Early stages in the development of speech movements. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 37-50.) New York: Stockton Press.

Kuhl P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69-72.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discrimination at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics, 32,* 542-550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003-1010.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception* (Vol. 2). New York: Academic Press.

Lack, D. (1961). *Darwin's finches: An essay on the general biology of evolution.* New York: Harper.

Lehmann, W. P. (1973). *Historical linguistics.* New York: Holt, Rinehart and Winston.

Lenneberg, E. H. (1967). *The biological foundations of language.* New York: John Wiley.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Lindblom, B. (1983). Can the models of evolutionary biology be applied to phonetic problems? In A. Cohen & M. P. R. Broecke (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences.* Dordrecht, Holland: Foris.

Lindblom, B. (1989). Some remarks on the origin of the "phonetic code." In C. von Euler, I. Lundberg, & G. Lennerstrand, (Eds.), *Brain and reading* (pp. 27-44). Basingstoke, England: MacMillan.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of language universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for language universals* (pp. 181-203). The Hague: Mouton.

Locke, J. (1983). *Phonological acquisition and change.* New York: Academic Press.

Locke, J., & Pearson, D. M. (1988). Linguistic significance of babbling: Evidence from a tracheostomized infant. *Report from the Neurolinguistics Laboratory at Massachusetts General Hospital* (Whole issue).

Locke, J. L., & Mather, P. L. (in press). Genetic factors in the ontogeny of spoken language: Evidence from monozygotic and dizygotic twins. *Journal of Child Language.*

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of organization. *Lingua, 49,* 11-49.

MacNeilage, P. F. (in press). The "postural origins" theory of primate neurobiological asymmetries. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayr, E. (1982). *The growth of biological thought.* Cambridge, MA: The Belknap Press.

McCune, L., & Vihman, M. (1987). Vocal motor schemes. *Papers and Reports on Child Language Development, 26,* 72-79

Menn, L. (1983). Development of articulatory, phonetic and phonological capabilities. In B. Butterworth (Ed.), *Language production* (Vol. II, pp. 3-50). London: Academic Press.

Newport, E. L., & Meier, R. P. (1985). The acquisition of American Sign Language. In D. I. Slobin (Ed.), *The cross-linguistic study of language acquisition, Volume I: The data* (pp. 881-938). Hillsdale, NJ: Lawrence Erlbaum Associates.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1: Production* (pp. 93-112). New York: Academic Press.

Oller, D. K. (1986). Metaphonology and infant vocalizations. In B. Lindblom, & R. Zetterström (Eds.), *Precursors of early speech* (pp. 21-35). New York: Stockton Press.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59,* 441-449.

Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech and Hearing Research, 28,* 47-63.

Oyama, S. (1985). *The ontogeny of information.* New York: Cambridge University Press.

Piaget, J. (1978). *Behavior and evolution.* Chicago: Chicago University Press.

Roeper, T. & Williams, E. (Eds.) (1987). *Parameter setting.* Dordrecht, Holland: D. Reidel.

Savage-Rumbaugh, E. S. (in press). Language learning in the Bonobo: How and why they learn. In N. Krasnegor, D.

Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109-136). New York: Springer-Verlag.

Stark, R. E. (1986). Prespeech segmental feature development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 149-173). New York: Cambridge University Press.

Stern, D. N. (1985). *The interpersonal world of the infant*. New York: Basic Books.

Stoel-Gammon, C., & Otomo, K. (1986). Babbling development of hearing-impaired and normally hearing subjects. *Journal of Speech and Hearing Disorders, 51*, 33-41

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 58-76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology, 17*, 237-257.

Thelen, E. (in press). Motor aspects of speech. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thompson, D. W. (1917/1961). *On growth and form* (abridged ed.). New York: Cambridge University Press.

Trevarthen, C., & Marwick, H. (1986). Signs of motivation for speech in infants, and the nature of a mother's support for development of language. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 279-308). New York: Stockton Press.

Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics, 7*, 3-40.

Vihman, M. M., & Miller, R. (1988). Words and babble at the threshold of lexical acquisition. In M. D. Smith & J. L. Locke (Eds.), *The emergent lexicon: The child's development of a linguistic vocabulary* (pp. 151-183). New York: Academic Press.

Waddington, C. H. (1957). *The strategy of the genes*. New York: MacMillan.

Waddington, C. H. (1975). *The evolution of an evolutionist*. Ithaca, NY: Cornell University Press.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Walley, A. C., Pisoni, D. B., & Aslin, R. N. (1981). The role of early experience in the development of speech perception. In R. N. Aslin, J. Alberts, & M. R. Petersen (Eds.), *Development of perception, Volume 1* (pp. 219-255). New York: Academic Press.

FOOTNOTES

[*] In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants*. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

[†] Also University of Connecticut and Yale University.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 28-40

Phonemic Segmentation as Epiphenomenon: Evidence from the History of Alphabetic Writing*

Alice Faber

"...we must not be misled in our ontology by the possibilities provided by our metalanguage."

(Anderson 1985: 9)

1. PRELIMINARY REMARKS

There is by now a large and convincing body of evidence that linguistic units representing acoustic or articulatory steady states need not be included as primitives in linguistic representations of phonological structure. Alternatives to such segments[1] being pursued in current phonological work include both larger units that in traditional analysis might be treated as more than one segment and single features that might be part of the representation of one or more segments in a word. Similarly, investigations of language use suggest that many speakers do not divide words into phonological segments unless they have received explicit instruction in such segmentation comparable to that involved in teaching an alphabetic writing system. Nevertheless, alphabetic writing, writing whose symbols represent individual segments, exists, and is learnable.

Paradoxically, then, alphabetic writing is based on a phonological unit that is arguably nota natural unit (see also Studdert-Kennedy, 1987: 68). How is this paradox to be resolved? To the extent that the paradox has been recognized, the segmental nature of alphabetic writing has been taken as paramount: Alphabetic writing could not have evolved if the segmentation on which it is based were not natural, and, therefore, the existence of alphabetic writing is itself evidence for the naturalness of phonological segmentation (Öhman, 1979: xviii; Saussure, 1959: 39; Studdert-Kennedy, 1987: 68).

In the absence of a convincing explanation for the innovation of alphabetic writing that does not rely on the prior existence of segmentation ability in the human repertoire, explanations of this nature, however dissatisfying, cannot be rejected.[2] My purpose in this paper is to provide just such an alternative. After a brief review of some of the evidence regarding segments and segmentation ability, I will provide a detailed exposition of how the Greek alphabet might have developed. On the basis of this scenario, I will suggest that segmentation ability, rather than being a necessary precursor to the innovation of alphabetic writing, was a consequence of that innovation.

2. THE STATUS OF PHONEMIC SEGMENTATION

As already noted, most twentieth century approaches to phonological analysis rely on the idealization that the speech stream is divided into discrete segments, each representing an acoustic or articulatory steady state. Nevertheless, as is well known, there is no physical basis for this segmentation.[3] As a consequence, there is a long and often under-appreciated tradition of alternative approaches to phenomena not easily amenable to segmental analysis, most notably of phenomena not easily localized within a word (e.g., Firth, 1957; Harris, 1944; Palmer, 1970).

I would like to thank Lloyd Anderson, Cathe Browman, Peter Daniels, Len Katz, Ignatius Mattingly, M. P. O'Connor, and Sue Schmerling for comments, discussion, and suggestions which, although not always incorporated herein, are nonetheless greatly appreciated. Extensive discussion with Tom Sawallis contributed to the early growth of my ideas on this topic. Support from NIH grant NS-007237 (to Haskins Laboratories) during the writing of this paper is gratefully acknowledged.

Within "mainstream" theoretical phonology in the decade following the publication of Chomsky and Halle (1968) (SPE), one of the major innovations was the integration of the syllable into linguistic models. This integration was prompted by the more elegant and perspicuous treatment of stress in the newer models (Hayes, 1980; Liberman & Prince, 1977; McCarthy, 1979) as well as by the range of apparently segmental phenomena conditioned by syllable position (Hooper, 1972; Kahn, 1976).[4] At least some of these alternatives refer also to constituents of the syllable (e.g., Halle & Vergnaud, 1980: 93; Selkirk, 1982; cf. Clements & Keyser, 1983 and Hooper, 1972):[5] the vocalic *nucleus*, the pre-nucleus consonantal *onset*, and the post-nucleus *coda*. The nucleus and the coda are generally grouped together as the *rhyme*. Just how these hierarchically arranged units should be related to previously hypothesized units like segments is not explicitly discussed. Rather, it is assumed that syllable constituents dominate segmental nodes, so that a hierarchical syllable display can be unambiguously related to a distinctive feature matrix of the customary sort (Clements & Keyser, 1983: 25), a column on such a matrix corresponding to a phonemic segment.

Given the demonstrable relevance of syllable components in phonological structure, maintenance of segments in syllable-based analyses is surprising. The rich inventory of English syllable types gives rise to a tacit feeling that the inventory of phonological primitives would be prohibitively large. To my knowledge, only Fujimura and Lovins (1978) have proposed an analysis based on demisyllables (approximately, onsets and rhymes) and extra-metrical consonantal affixes. As Fujimura (1980: 122) notes, the memory burdens imposed by an inventory containing c. 1000 elements would not be excessive; Browman's (1980) demi-syllable-based speech synthesis program requires c. 850 demi-syllables and affixes. This increased inventory size (relative to traditional analyses) is compensated for by a greatly decreased need for contextually conditioned phonological rules; Browman's model, for example, requires 20 rules.

One of the major European contributions to phonological theory has been an emphasis on the paradigmatic relationships among the segments of a language. Thus, for Trubetzkoy (1949), a phoneme is defined by its contrasts with other phonemes. From here it is a short step to the distinctive features introduced by Jakobson, Fant and Halle (1969), a formalization of the dimensions along which segments can differ or be similar. While the Jakobsonian distinctive features were based almost exclusively on acoustic properties of segments, virtually all current feature-based models rely on articulatory features.

In SPE, distinctive feature matrices were simply a two-dimensional array of rows representing features and columns representing segments. All of the specifications for one segment preceded any of the specifications for the next. This two-dimensional array has evolved in descendants of SPE, so that the rows are spread out in three dimensions (Clements, 1985; Sagey, 1986; Schein & Steriade, 1986). And, given underspecification of feature values (Archangeli, 1985), the notion of a column representing an individual segment is greatly obscured; individual feature specifications may, in underlying representation, precede other feature specifications, but, without the root node (Sagey, 1986: 40–44, 275), to which all specifications for a particular segment are linked, it would be impossible to make precedence statements about segments independent of their featural constituents.

Statements about the syntagmatic relationships among individual features, independent of their putative segmental affiliation, are an integral part of the model of articulatory phonology being developed by Browman and Goldstein (1986, 1987, in press). In this model, the primitives are articulatory gestures, represented on a gestural score, a two-dimensional matrix. In contrast to the [±F] values of more traditional distinctive feature models, gestures are either present or absent, and, as already noted, the temporal relationships among gestures on the same score are not constrained, but must be specified on a language specific, or even a style specific, basis.

The approaches just sketched, converging on phonological representations in which segments *per se* play little, if any, role, also allow for more perspicuous treatment of a class of sounds that has been chronically insusceptible of principled treatment in segment-based models—diphthongs, affricates, and prenasalized consonants. These are sounds that, by their very nature, cannot be conceived of as single steady states. As a result of their inherently dynamic nature, their existence has given rise to long and, for the most part, fruitless discussions about whether each is best treated as "one segment or two," as a complex unit or a simple cluster.[6]

To summarize, there is little unambiguous structural support for positing segments as linguistic units. In particular, there is no need to appeal to segments in modeling ordinary language structure and use. Neither is there evidence that incorporation of segments into linguistic models leads to more satisfying or parsimonious analyses. In contrast, structural linguistic evidence suggests that models based on syllables and syllable components (either onset-rhyme or onset-nucleus-coda) might be more appropriate. However, there are spheres, notably reading, in which language users clearly do display an ability to recognize and manipulate segments.

The English alphabetic writing system encodes, albeit not always systematically, the segmental phonemic structure of English. While skilled readers may not always make direct use of the phonological coding inherent in an orthographic form, they clearly can do so in the case of rare or unknown words. Studies of the development of reading suggest that at least a minimal awareness of segmentation is necessary for English speaking children to learn to read (Mattingly, 1972: 44). Small children tend to be much more proficient at syllable counting tasks than at phoneme (segment) counting tasks, while literate older children and adults can do both (Liberman & Shankweiler, 1987: 207).

Given a correlation between segmental awareness and literacy (in a segment-based alphabet), the question arises whether segmentation ability, arising spontaneously as part of cognitive maturation, is an indication of reading readiness, or whether segmentation ability arises as a consequence of specific teaching. Liberman and Shankweiler (1987: 210) refer specifically to reading instruction, but widely available alphabet-related toys and books could easily contribute to incipient segmental awareness on the part of pre-school children. Nevertheless, children who are aware of the alphabet but who cannot yet read may interpret alphabetic symbols differently than do proficient readers. Read (1986: 51, 105) refers to two relevant types of studies. In one, conducted in Argentina, preliterate children interpreted each letter as representing a syllable. In another pair of studies, conducted in the United States, some beginning spellers were found to write consonants only, omitting the vowels.

Further relevant findings concern adult populations. Morais et al. (1986) studied adult illiterates and students in adult literacy classes in Portugal. Newly literate adults outperformed illiterates on tasks involving phonemic segmentation, but not on comparable tasks involving non-linguistic segmentation (i.e., of notes in a melody). Similarly, Read et al. (1986) compared two groups of literate Chinese, those who had never been exposed to the *pinyin* Romanization, and those who had, even if they were no longer proficient in it. The *pinyin* group outperformed the non-*pinyin* group in tasks involving segment addition or deletion. Likewise, Mann (1986) found that Japanese elementary school children are less proficient than their American counterparts at performing segment-related tasks.[7] One basic conclusion can be drawn from these studies: segmental awareness is a result not of cognitive maturation but of exposure to a segment-based orthography; literacy itself is not sufficient.

Reading is clearly a metalinguistic ability. There are languages for which there exist neither a standard orthography nor a written literary tradition. Even some languages with long literary traditions (e.g., Amharic) have large numbers of illiterate speakers. Furthermore, the whole question of 'functional' literacy suggests that there exist within literate populations substantial differences in reading proficiency. Thus reading competence cannot be equated with language competence in general. If segmentation ability results primarily (if not only) from exposure to segmental orthography, it too must be treated as a metalinguistic ability, and thus it cannot be ascribed to general linguistic competence.[8]

Language games have been observed in many cultures, and have often been seen as sources of evidence regarding linguistic structure. Language games generally involve systematic insertion of extraneous linguistic material and/or permutation of structural linguistic units. Some games are best defined on syllable components, while others are best defined on segments; in games defined on syllable units, initial clusters like *sp-* would act as a unit, while in games defined on segments, *s* and *p* could act independently. Bagemihl (1987), in an extensive survey of documented language games, reports that games based on syllables are far more common than games based on segments; the latter naturally presuppose segmental awareness on the part of their speakers. He further notes (1987: 36) that segment-based games are attested only in languages with alphabetic writing systems, suggesting that segmental awareness here too is dependent on exposure to alphabetic writing.

3. ALPHABETIC WRITING

The evidence just summarized suggests that, despite their intuitive appeal and longevity, segment-based analyses might not be optimally suited to represent language as a cognitive system. A careful examination of the complex relationship between orthographic and linguistic units is thus in order. This examination will proceed in two stages. First, I will motivate a typology of writing systems. Then, on the basis of this typology, I will discuss the extent to which the development of orthographic systems bears on the question of phonemic segmentation, and just what that bearing is. My goal is, in particular, to trace the development of alphabetic writing, since it is the existence of alphabetic writing that is used (implicitly or explicitly) as evidence for the universality of segments as a building block of language. My discussion is informed by the view expressed by O'Connor (1983: 441) that the structure of an orthography for a particular language reflects, albeit not always systematically, native speaker analysis of that language. A script created for a language *ex nihilo* will reflect both the level of appropriate units and the inventory of units at that level. In contrast, a borrowed script may inherit the level of units, but still reflect in some measure the inventory of the borrowing language.

3.1. Typology of writing systems

An *orthography* can be defined as a system of markings which can unambiguously cue a trained native speaker to produce acceptable utterances in that language (Faber, 1987). Similar definitions have been offered by Sampson (1985: 19) ("a given set of written marks together with a particular set of conventions for their use") and Daniels (1986: 1) ("the sets of marks by which utterances are notated so as to be precisely reproducible in the absence of the notator or the notator's instructions"). These definitions agree in referring to the arbitrary yet non-idiosyncratic nature of the signs and to the training involved in their interpretation, as well as in not restricting the linguistic level of the units. My definition further excludes notational systems that can be interpreted in arbitrarily many ways by speakers of the same language (or, for that matter, of different languages). Thus, the set of "iconic" international road signs does not represent an orthography, despite the relatively arbitrary nature of the symbols and their supposedly unambiguous semantics.

3.1.1. Terminology

Definition of orthography provides a useful point of departure for discussion of orthographic systems. But it is not enough. Before attested orthographies can be classified in any meaningful typology, it is necessary to define both the dimensions along which they can differ and the resultant categories. Not to do so invites confusion. The debate over whether the 2nd. and 1st. millennium BC West Semitic scripts—Canaanite, Ugaritic, Phoenician, Aramaic, Old South Arabian, in plain type in Figure 1,[9] but not their vocalized descendants, in bold type—are syllabaries (Gelb, 1963; Swiggers, 1983, 1984), alphabets (Daniels, 1986; Naveh, 1982), or neither (as I will argue shortly) hinges not on differing conceptions of the West Semitic scripts but rather on differing definitions of alphabet and syllabary. Except where noted, the following definitions are based on Sampson (1985: 32ff).

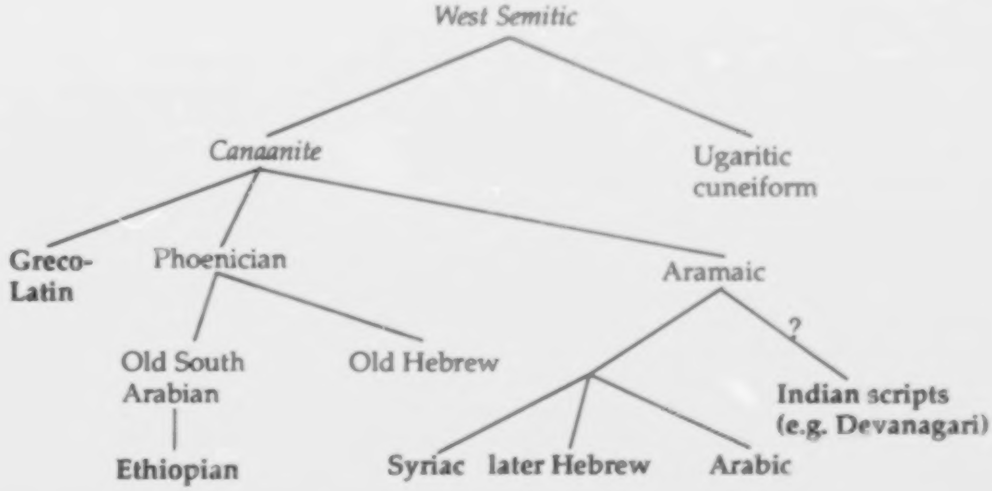


Figure 1. (Simplified) depiction of relationships among Semitic orthographies. Legend: italic—hypothesized, unattested scripts; plain—scripts representing primarily (or only) consonants; bold—scripts representing consonants and vowels.

Logographic vs. Phonographic. Orthographies differ in whether the bulk of the symbols in the system, taken in isolation, are susceptible of semantic interpretation. The English symbol <d> has no inherent meaning, while the symbol <+> does. Systems in which symbols like the former predominate are *phonographic* while systems in which symbols like the latter predominate are *logographic*. Logographic systems code for the most part morphemes, as Sampson notes, and phonographic systems code units of sound.

Complete vs. Defective. An additional dimension along which orthographies vary is the extent to which they aspire to code all units of the relevant level. *Complete* orthographies code all (or virtually all) relevant units, while *defective* or *incomplete* orthographies omit some. Orthographies can fail to be complete in two distinct ways. The primary way in which orthographies can be defective is that certain units may simply not be represented. This sort of omission is the basis for the popular misstatement that Hebrew at one time didn't have any vowels. The earliest Hebrew script, like the other West Semitic scripts already referred to, did not contain any symbols for vowel sounds, even though the language had a typologically normal inventory of vowels. Another way in which orthographies may be incomplete (one which will not be discussed further here) is that one or more phonemic contrasts may not be represented, the same symbol being used indifferently to represent two or more phonologically distinct units. Thus, the English digraph <th> is used for both voiced and voiceless interdental fricatives. In addition, it is fairly common for orthographies that are otherwise fairly complete not to indicate quantity and tonal contrasts, even in languages in which such contrasts bear a heavy functional load. As Sampson notes (p. 36), no orthography is ever truly complete. Most orthographies notate few or no non-contrastive features; indeed special purpose orthographies that do aspire to this level of completeness (e.g., narrow phonetic transcription) are notoriously difficult even for trained individuals to read fluently.

Syllabic vs. Segmental. Phonographic scripts differ with regard to the linguistic level that they code. Thus symbols in syllabic orthographies code syllables, and symbols in segmental orthographies code segments.[10] It is, I believe, plausible to claim that the syllable is a relevant linguistic unit for users of a syllabic orthography. But the parallel claim regarding segmental orthographies is less plausible. Certainly, users of a complete segmental orthography must, at some level, be aware of those units represented by its symbols. But, as noted in the previous section, some orthographies are systematically incomplete. And it does not follow that users of such defective orthographies are aware of units or contrasts that are not reflected in their orthography. In fact, the appropriate implication may well be in the reverse direction; the failure of the early Semitic segmental orthographies to code vowels may indicate that users of these orthographies did not segment syllables like /ba/ into their constituent parts, despite their awareness of the similarity among /ba/, /bu/, /bi/, and even perhaps /tab/. Awareness of this similarity, while *prerequisite* to segmentation, does not, as noted above, *imply* exhaustive segmentation of the speech stream (cf. Mattingly, 1987). Rather, the vowel segments not coded in these orthographies constitute the background for the consonantal segments, which are coded.[11]

The question is not, and *cannot* be, whether users of early Semitic consonantal orthographies had the ability to recognize and manipulate the phonemic segments of their language. It is, rather, which abilities can be imputed to users of these scripts on the basis of the structure of the scripts. Segmentation ability presupposes an awareness not only of the similarity among /ba/, /bi/, and /tab/, but also of the similarity among /ba/, /ta/, and /ʔa/. It further implies localization of these similarities in the segments /b/ and /a/, respectively. Evidence that users of an orthography had one of these abilities (in this case, the first), does not speak to whether they had either (or both) of the other two; they may have, or they may not.

Linear vs. Non-linear. Orthographies also differ in regard to linearity. In a *linear* orthography, the order of the symbols, either horizontally or vertically, iconically mirrors the order of the sounds that they represent. Thus, in "phonetic languages" like Spanish, the letters in sequence reflect the sequence of sounds. In contrast, English orthography is not as linear. The contrast in the medial vowels of *cop* /kap/ and *cope* /kop/ is orthographically conveyed by the presence of final *–e* in the latter, so the vowel /o/ is represented by the discontinuous sequence of letters *o–e*. In more systematically non-linear scripts like Egyptian hieroglyphics or Korean Han-gul (Sampson, 1985: 131), the placement of symbols may be governed by aesthetic rather than by linguistic principles. Other major non-linear scripts include two segmental scripts that are commonly and erroneously treated as syllabaries: Indian Devanagari (and allied South Asian scripts) and the Ethiopian script. While the organizational pattern of both is more or less syllabic, as Sampson (1985: 66) notes, the individual graphs clearly encode segmental information. However, as shown in Figure 2,[12] the placement of the graphic modifications for vowel content in the Ethiopian script, represented here by Amharic, is such that a CV syllable may, depending on the vowel, be modified on either the left side or the right side of the symbol.[13]

Even those who recognize the linearity factor assume that the linearity of a phonographic script is less significant than the linguistic level encoded in its symbols. That is, the set of syllabic scripts *and* the set of segmental scripts can both be divided into linear and non-linear subsets. I would now like to propose an alternative view. In that view, linearity is hierarchically superordinate to the syllabic/segmental dimension. Furthermore, the linguistic level at which a script can be seen as linear is logically distinct from the linguistic level of the units that it encodes. Thus, the Ethiopian script encodes segments but is syllabically linear, while European scripts encode segments and are segmentally linear.

This property of the Ethiopian scripts is illustrated in Figure 3 with the Amharic word /leba/ "thief." Note the non-linear mapping of graphs to segments, contrasted with the linear mapping of graphs to syllables. Of course, any segmentally linear script will necessarily be syllabically linear, while the reverse is not true.[14]

Alphabet. An *alphabet* can be defined as a script that is (relatively) complete and, in its ideal case, segmentally linear. By this definition, the early Semitic scripts, which encode for the most part consonants, are not alphabets; since they do not encode vowels, they are systematically incomplete. Likewise, the later Semitic systems that, like Arabic, do encode some vowels, are not, strictly speaking, alphabets, since they are syllabically rather than segmentally linear.[15] The diacritic markings that indicate vowel quality in Arabic, Hebrew, and Syriac may occur above, below, within or after (to the left of) the symbol for the preceding consonant sound.[16] I am, in effect, giving special definitional status to just those orthographies whose existence has been taken to reflect the innate naturalness of phonemic segmentation. The orthographies that I define as alphabets turn out to be those that have equally salient symbols for vowels and consonants. An English speaker unfamiliar with Russian will have no *a priori* intuitions about whether a particular Russian symbol represents a vowel or a consonant. One might say, then, that an alphabet is a script which treats vowels and consonants equally and that, in order to do so, a script must paradoxically recognize vowels and consonants both as separate and equal.[17]

	ä (ə)	u	i	a	e	ə (ɨ)/0	o
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
w	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
y	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
ṭ	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ

Figure 2. A subset of the Amharic orthography. Amharic, a Semitic language, is the national language of Ethiopia.

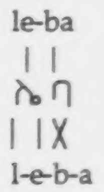


Figure 3. **The segmentally coded, syllabically linear nature of the Ethiopian script, illustrated with the Amharic word LEBA "thief."**

In contrast, Daniels' (1986: 8) definition of an alphabet as a system in which "each character conveys information about just one segment"[18] implies that the primary subdivision of segmental scripts should be based not on properties of the sound-to-symbol mapping but rather on the visual autonomy of the symbols participating in it. Naveh's (1982: 11) functional definition of an alphabet as an orthography containing a limited number of symbols (20-30) that can be arranged in a fixed citation order (A B C D ...), which, like Daniels', includes the West Semitic scripts, is, I believe, inappropriate, in that it makes explicit reference to metalinguistic, perhaps pedagogical, facts rather than to the structure of the sound-symbol mapping. It should be noted that my disagreement with Daniels and Naveh concerns *only* whether the term "alphabet" should be restricted to complete, segmentally linear orthographies. I share Daniels' view that the Greek development of a complete, segmentally linear orthography, by whatever name, is not to be causally linked to the development of Greek, and hence Western, civilization. And, I most emphatically would not follow Gelb and, perhaps, Swiggers in treating those segmentally linear Semitic scripts that do not encode vowels as syllabaries. It is not an "either-or" question. The dimensions phonographic/logographic, syllabically/segmentally linear, complete/defective, and syllabically/segmentally encoded define five categories of phonographic orthographies, as illustrated in Figure 4.

My focus on which of these categories contain orthographies that it is appropriate to label "alphabet" has no bearing on how orthographies in the other categories should be labeled.[19] Interestingly, Sampson (1985) offers no definition of alphabet, but uses the term at various points to refer to the Northwest Semitic scripts. In any case, the set of alphabets is clearly a subset of the set of segmental orthographies.

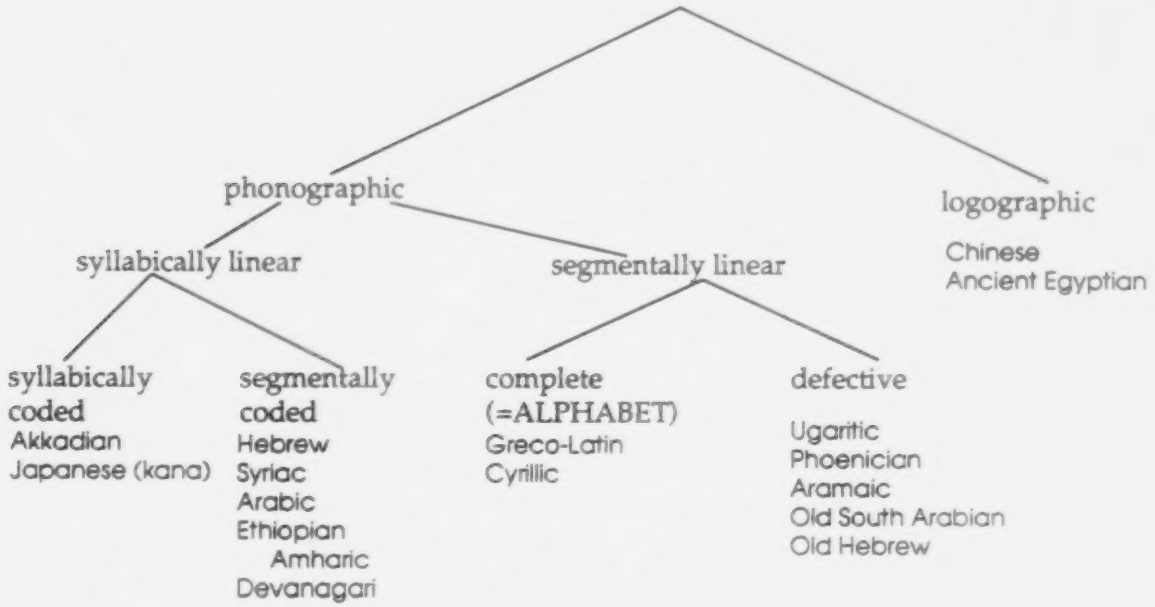


Figure 4. Some representative orthographies arranged according to the types defined in the text.

3.1.2. Distribution of writing systems

If the distribution of orthographic types is charted based on the number of users, alphabetic writing systems are by no means the most widespread. While alphabets are used by about a billion literate individuals,[20] world-wide, a comparable number use either logographic writing (Chinese) or syllabically linear systems, used in Northeast Africa (Ethiopian syllabary), Southern Asia (Devanagari and allied systems), and Japan. If the Arabic-based scripts (used for Farsi, Urdu, Pashto, Malay, etc., in addition to Arabic) which, as already noted, are syllabically linear when vocalized even though they encode segments, are included, the distribution of syllabically linear scripts is even greater. Thus, (*pace* Lotz, 1972: 119) syllabically linear scripts (which he refers to as syllabaries) are not typologically marginal; both the Japanese true syllabary and the syllabically linear segmental scripts mentioned above have been quite stable for 1500 years and more.

The Greco-Latin scripts, including Cyrillic, while geopolitically of great importance, are more circumscribed demographically. And most of their expansion has occurred in the past 500 years. In the more distant past, the relative importance of syllabically linear scripts was much greater. The further back one goes in the history of writing, the greater the demographic preponderance of non-linear and syllabically linear scripts: Sumero-Akkadian cuneiform, and Linear B, used for Mycenean Greek,[21] were syllabically linear and incomplete; Egyptian hieroglyphic writing was, as already noted, non-linear.

Given the five-thousand year long, incompletely documented history of writing it is tempting, and, perhaps, natural to see, as does Gelb, a developmental, culturally maturational trend toward segmentally linear scripts—that is, toward alphabets as defined earlier. However, such a view would be inappropriate, not to mention ethnocentric. All known complete, segmentally coded orthographies (see Figure 4) arose from the early West Semitic segmentally linear incomplete scripts (Sampson, 1985: 77). All of these complete descendants maintained linearity—the Greco-Latin scripts on a segmental level, and the vocalized Semitic scripts and the Ethiopian and Devanagari scripts (if Devanagari is, in fact, to be traced to West Semitic origins) on a syllabic level. Furthermore, the West Semitic segmentally linear incomplete script is the only known representative of its type. In order to attribute the innovation of either a segmentally linear script or of an alphabet to cultural maturation, we would need evidence of a recurrent pattern. While such evidence is not *in principle* impossible, it is, possibly as a result of an accident of human history, the case that each of these innovations is unique. It is thus *in practice* impossible to determine whether either innovation was a sign of cultural progress or merely the product of the special circumstances in which it took place. To the extent that explanations for these innovations can be based on the specific circumstances in which they took place, it is unnecessary, unparsimonious, and even imprudent to hypothesize more abstract, putatively causal principles.

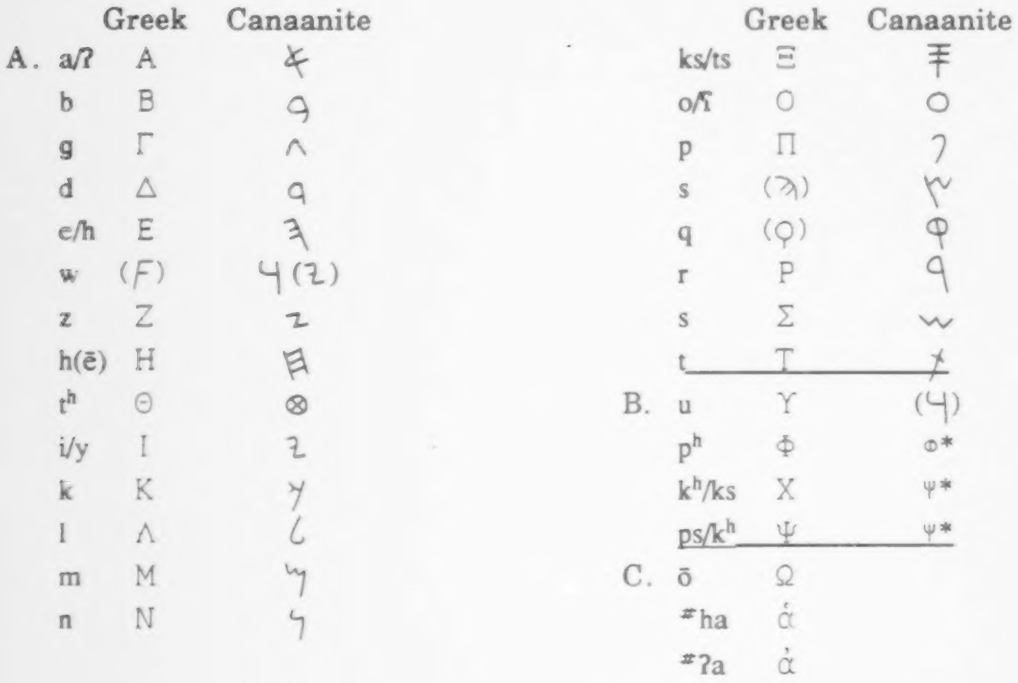
3.2. The Greek adaptation of the Canaanite system

If it is only use of an alphabet that bespeaks awareness of phonemic segmentation, then it is worth examining the early development of the only alphabet ever known to have arisen spontaneously in human history, the Greek alphabet. The Classical Greek orthography of c. 400 BC, illustrated in Figure 5, has seventeen symbols for consonants and seven for vowels. Several phonemic contrasts (/ŋ/ vs. /g/,[22] /a/ vs. /ā/, /i/ vs. /ī/, /ü/ vs. /ǖ/) are not directly represented, but, aside from these marginal 'defects', comparable to those observed in later adaptations of the orthography, the Greek alphabet departs from the ideal of a complete, segmentally linear orthography in two respects: (1) the rough and smooth breathing marks, post-classical innovations which denote /h/ and /ʔ/ onsets to an initial vowel, are written as diacritics *over* the letter representing the vowel: ἁ, ἀ. And, (2) there are symbols for two clusters /ks/ and /ps/, which, as Sampson (1985: 103) notes, following Allen (1974: 57), are the only two clusters to occur in coda position in ancient Greek.[23] The letters Ξ *ksi* and Ψ *psi* are used, however, even when the clusters are split by a morpheme boundary. The orthography of the earliest Greek inscriptions, which date from the eighth century BC, while archaic in some respects, is typologically comparable to that of later Greek, that is to say, alphabetic.

The accepted wisdom, in, e.g., Jeffery (1961), is that the Greeks adopted the Phoenician alphabet in the eighth century BC, shortly before the oldest attested inscriptions. However, in light of the vast increase in recent years in the size of the Greek and Phoenician inscriptional corpora (as summarized in, e.g., McCarter, 1975), as well as increased sophistication in dating inscriptions in all relevant languages, the situation is less clear. Isserlin (1983) notes a lack of scholarly consensus regarding the time of the borrowing (eighth century BC, as argued by Jeffery, or eleventh century BC, as suggested by Naveh [1972]), the geographical context of the borrowing (Aegean, Levant, or Asia Minor), and the orthographic source of the Greek alphabet (Phoenician, Aramaic, or (unattested) Canaanite; Kaufman, 1987). As all of these potential sources were typologically comparable in all relevant respects, I will, for present purposes, refer only to Canaanite.

Any explanation for the development of the Greek orthography must take into account that the development was essentially complete by the time of the earliest Greek inscriptions; there are no traces available either of Canaanite (Phoenician or Aramaic) representation of vowels comparable to that attested in Greek[24] or of Greek material in which vowels are not represented. Furthermore, all of the symbols which represent Greek vowels (with the trivial exception of the late Ω) are derived from Canaanite letters for sounds that were, at best, marginal in Greek: /ʔ h ʕ y w/. While it is, indeed, possible that these symbols were deemed superfluous by the importers of the Canaanite system, and thus used for sounds not represented by that system, it is not necessary to base an explanation for the structure of the Greek alphabet on the unattested existence of an unknown genius.

The names of the Greek letters *alpha, beta*, etc., meaningless in Greek, have clear sources in a Canaanite acrophonic tradition, whereby each sound is associated with an object whose name begins with that sound. This fixed order of a traditional, invariant list is comparable to modern radio alphabets like *able, baker, charlie*, etc.



		Greek	Canaanite
A.	a/ʔ	A	𐤀
	b	B	𐤁
	g	Γ	𐤂
	d	Δ	𐤃
	e/h	E	𐤄
	w	(Ϝ)	𐤅 (𐤆)
	z	Z	𐤆
	h(ē)	H	𐤇
	t^h	Θ	𐤈
	i/y	I	𐤉
	k	K	𐤊
	l	Λ	𐤋
	m	M	𐤌
	n	N	𐤍

		Greek	Canaanite
	ks/ts	Ξ	𐤎
	o/ʕ	O	𐤏
	p	Π	𐤐
	s	(Ϻ)	𐤑
	q	(Ϙ)	𐤒
	r	P	𐤓
	s	Σ	𐤔
	t	T	𐤕
B.	u	Υ	(𐤅)
	p^h	Φ	Φ*
	k^h/ks	X	Ψ*
	ps/k^h	Ψ	Ψ*
C.	ō	Ω	
	#ha	ἁ	
	#ʔa	ἀ	

Figure 5. The Greek alphabet with its Canaanite precursor, represented by an Old Aramaic script. Parenthesized Greek letters are found in archaic materials only, and are not preserved in Classical Greek. Parenthesized Canaanite symbols represent deviations from the Semitic citation order. Group A includes Greek letters adopted directly from a Canaanite source and mirroring the source citation order. Where two values for these letters are given, the first one represents the Greek value and the second the Semitic value. Group B contains letters based on Semitic prototypes that were added to the end of the Greek alphabet. The alternate values represent variation in early Greek orthographic traditions. The starred source symbols are potential South Arabian symbols for Greek letters which have no obvious Canaanite prototype. Group C represents indigenous Greek developments.

As Sampson (1985: 101) demonstrates, conclusively to my mind, transmission of the Canaanite script, using the acrophonic principle would have led to the *mis*interpretation of several Canaanite consonant symbols as representing vowels instead. The Canaanite words *ʔalpa* "cow," *he* "?," *yoda* "hand," and *ʕayna* "eye," standing for 𐤀/ʔ/, 𐤄/h/, 𐤉/y/, and 𐤏 /ʕ/, would have been perceived by speakers of a language in which, as in Greek, these sounds did not occur, as beginning in [a], [e], [i], and [ɑ], respectively.[25] Thus, Phoenician [ʔalpa], with an initial [ʔ] became Greek [alpa], with no [ʔ]. Identification of [a] and [ɑ] with Greek /a/ and /o/ is plausible,[26] and the use of 𐤉 *yod* for /i/, not implausible from a phonetic point of view, may be related as well to the near absence of /y/ in Greek and to early Canaanite scribal practices like those observed for Ugaritic, in which <y> occasionally represented /i/. The form of Υ *upsilon* appears to be based on the archaic Canaanite 𐤅 *wau*, whose place in the early Greek alphabetical order was filled by Ϝ *digamma*, the shape of which is in turn based on Canaanite 𐤉 *yod* (Allen, 1974: 45).[27]

4. IMPLICATIONS

In this paper, I have demonstrated that it is necessary neither to impute to users of West Semitic consonantal scripts segmentation ability, as commonly understood in psycholinguistics research, nor to appeal to such ability to explain the Greek innovation of alphabetic writing. In fact, segmentation ability as a human skill may have been a direct result of (rather than an impetus to) the Greek development of alphabetic writing. Thus, the existence of alphabetic writing can not be taken *eo ipso* as evidence for the cognitive naturalness of the segmentation that it reflects. Given this conclusion, it is necessary to pose the question of why virtually all linguists have fallen prey to what Ladefoged (1984: 92) has, in a comparable context, referred to as a phonemic conspiracy. That is, we as linguists feel that, because we *can* describe linguistic systems in terms of phonemic segments, we *must* do so. That we are influenced in this practice by the structure of the alphabets that we use in our ordinary lives is obvious. As far as I know, every technical linguistic tradition that refers to segments arose in an alphabetic milieu or was influenced directly by such a tradition. There is no segmental analytic tradition not supported in this way by an orthographic tradition. In contrast, the indigenous Chinese linguistic tradition, described by Halliday (1981), has as phonological primitives syllable *initials* and *finals*, that is, onsets and rhymes. This analytical division is not supported by the logographic Chinese orthography, a lack which strengthens the force of the analysis.

Even though segment-based linguistic models are accurate in a wide range of cases, in particular in those in which syllable-constituent models reduce to segment models (e.g., CV syllables), the evidence that I have summarized in this paper suggests, however, that they are systematically limited in the extent to which they can model a wide range of linguistic data. These results lead to a challenge for future research, to develop and refine a rich set of non-segmental metaphors and notations, so that models based, like those sketched in Section 1 above, on different phonological primitives can be tested against a wide range of linguistic data, and revised appropriately.

REFERENCES

Allen, W. S. (1974). *Vox Graeca* (2nd. ed.). Cambridge: Cambridge University Press.

Anderson, S. R. (1974). *The organization of phonology*. New York: Academic Press.

Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. *Language, 52*, 326–344.

Anderson, S. R. (1985). *Phonology in the twentieth century*. Chicago: University of Chicago Press.

Archangeli, D. (1985). Yokuts harmony: Coplanar representation in nonlinear phonology. *Linguistic Inquiry, 16*, 335–372.

Bagemihl, B. (1987). *The crossing constraint and 'backwards' languages*. Manuscript, University of British Columbia.

Bange, L. A. (1961). *A study in the use of vowel letters in alphabetic consonantal writing*. Munich: UNI-Verlag.

Browman, C. P. (1980). Rules for demi-syllable synthesis using lingua, a language interpreter. *Proceedings of the ICASSP (IEEE)*.

Browman, C. P., & Goldstein, L. M. (1986). Towards an Articulatory Phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C. P., & Goldstein, L. M. (1987). Tiers in articulatory phonology, with some implications for casual speech. Haskins Laboratories *Status Reports on Speech Research, SR-92*, 1–30.

Browman, C. P., & Goldstein, L. M. (in press). Gestural structures and phonological patterns. To appear in I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chadwick, J. (1967). *The decipherment of linear B*. Cambridge: Cambridge University Press.

Chikanza, J. (1986). *Borrowing in Shona*. MA thesis, University of Florida.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. (1985). The geometry of phonological features. *Phonology Yearbook, 2*, 225–252.

Clements, G., & Keyser, S. J. (1983). *CV phonology*. Cambridge: MIT Press.

Cohen, G. L. (1982). The origin of the letter Omicron. *Kadmos, 21*, 122–124.

Daniels, P. T. (1986). Toward the linguistic study of writing: Aramaic orthographies. Annual Meeting of the Linguistic Society of America, New York.

Davis, S. (1987). On the arguments for syllable-internal structure. *Research on Speech Perception, Progress Report, 13*, 265–275.

Faber, A. (1985). Akkadian evidence for proto-Semitic affricates. *Journal of Cuneiform Studies, 37*, 101–107.

Faber, A. (1987). Phonetic and Phonological Interpretation of Orthographic Forms. Workshop on Language Change and Reconstruction Methodology, LSA Summer Institute.

Faber, A. (to appear). *Semitic sibilants: A study in comparative lexicography*. To be published by Eisenbrauns, Winona Lake, IN.

Firth, J. R. (1957/1948). Sounds and Prosodies. *Papers in Linguistics, 1934–1951*, 121–138. London: Oxford University Press.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguistics, 23*, 359–377.

Fujimura, O. (1980). Modern methods of investigation in speech production. In E. Fischer-Jørgensen & N. Thorsen (Eds.), *Proceedings of the the Ninth International Congress of Phonetic Science* (Vol. 3. pp. 117–135). Copenhagen: Institute of Phonetics.

Fujimura, O., & Lovins, J. (1978). Syllables as concatenative phonetic units. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 107–120). Amsterdam: North Holland.

Gelb, I. J. (1963). *A study of writing* (2nd. ed.). Chicago: University of Chicago Press.

Gordon, C. H. (1980). *Forgotten scripts*. New York: Basic Books.

Halle, M. (1964). On the bases of phonology. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language* (pp. 324–333). Englewood Cliffs, NJ: Prentice Hall.

Halle, M., & Vergnaud, J-R. (1980). Three dimensional phonology. *Journal of Linguistic Research, 1*, 83–105.

Halliday, M. A. K. (1981). The origin and early development of Chinese phonological theory. In R. E. Asher & E. J. A. Henderson (Eds.), *Towards a history of phonetics* (pp. 123–140). Edinburgh: University of Edinburgh Press.

Harris, Z. S. (1939). *Development of the Canaanite dialects*. New Haven: American Oriental Society.

Harris, Z. S. (1944). Simultaneous components in phonology. *Language, 20*, 181–205.

Hayes, B. (1980). *A metrical theory of stress rules*. Bloomington: Indiana University Linguistics Club.

Hayward, R. J., & Hassan, M. (1981). The Oromo orthography of Shaikh Bakri Sakalo. *Bulletin of the School of Oriental and African Studies, 44*, 550–566.

Herbert, R. K. (1986). *Language universals, markedness theory, and natural phonetic processes*. Berlin: Mouton de Gruyter.

Hooper, J. B. (1972). The syllable in phonological theory. *Language, 48*, 535–540.

van der Hulst, H., & Smith, N. (1982). An overview of autosegmental and metrical phonology. In H. van der Hulst & N. Smith (Ed.), *The structure of phonological representations*, (Vol. I. pp. 1–46). Dordrecht: Foris.

Isserlin, B. S. J. (1983). The antiquity of the Greek alphabet. *Kadmos, 22*, 151–163.

Jakobson, R., Fant, C. G. M., & Halle, M. (1969/1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jeffery, L. H. (1961). The local scripts of archaic Greece. Oxford: Clarendon Press.

Kahn, D. (1976). Syllable-based generalizations in English phonology. Bloomington: Indiana University Linguistics Club.

Kaufman, S. A. (1985). The pitfalls of typology: On the early history of the alphabet. *Hebrew Union College Annual, 57*, 1–15.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). *Speech: Articulation and perception*, tr. from the Russian. Washington: U.S. Department of Commerce.

Ladefoged, P. A. (1984). Out of chaos comes order: Physical, biological, and structural patterns in phonetics. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 83–95). Dordrecht: Foris.

Liberman, I. Y., & Shankweiler, D. (1987). Phonology and the problems of learning to read and write. *Memory and Learning Disabilities, Advances in Learning and Behavioral Disorders, Suppl. 2*, 203–224.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistics, 8*, 249–336.

Lotz, J. (1972). How language is conveyed by script. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 117–124). Cambridge, MA: MIT Press.

Mann, V. H. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*, 65-92.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 133–147). Cambridge, MA: MIT Press.

Mattingly, I. G. (1987). Morphological structure and segmental awareness. Haskins Laboratories *Status Report on Speech Research, SR-92*, 107–111.

McCarter, P. K. (1975). *The antiquity of the Greek alphabet and the early Phoenician scripts*. Missoula, MT: Scholars Press.

McCarthy, J. J. (1979). On stress and syllabification. *Linguistic Inquiry, 10*, 443–465.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12*, 373–418.

Mohanan, K. P. (1986). *The theory of lexical phonology*. Dordrecht: Reidel.

Morag, S. (1972). *The vocalization systems of Arabic, Hebrew, and Aramaic: Their phonetic and phonemic principles*. The Hague: Mouton.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*, 45–64.

Naveh, J. (1982). *Early history of the alphabet*. Jerusalem: Magnes Press.

O'Connor, M. P. (1983). Writing systems, native speaker analysis, and the earliest stages of Northwest Semitic orthography. In C. L. Myers & M. P. O'Connor (Eds.), *The word of the Lord shall go forth* (pp. 439–465). Winona Lake, IN: Eisenbrauns.

Öhman, S. (1979). Introduction. In Björn Lindblom & Sven Öhman (Eds.), *Frontiers of speech communication research* (pp. xv–xxv). London: Academic Press.

Packard, D. W. (1974). *Minoan linear A*. Berkeley: University of California Press.

Palmer, F. R. (1973) [1970]. Prosodic analysis. In E. C. Fudge (Ed.), *Phonology* (pp. 181–189). London: Penguin Books.

Read, C. (1986). *Children's creative spellings*. London: Routledge & Kegan Paul.

Read, C., Yun-Fei, Z., Hong-Yin, N., & Bao-Qing, D. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31–44.

Sagey, E. C. (1986). *The representation of features and relations in non-linear phonology*. Doctoral dissertation, MIT.

Sampson, G. (1985). *Writing systems*. Stanford: Stanford University Press.

de Saussure, F. (1959). *Course in general linguistics*. C. Bally & A. Sechehaye (Eds.), (Wade Baskin, Trans.) New York: McGraw Hill.

Schein, B. & Steriade, D. (1986). On geminates. *Linguistic Inquiry, 17*, 691–744.

Selkirk, E. O. (1982). *The syllable*. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. II, pp. 337–383). Dordrecht: Foris.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Schreerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

Swiggers, P. (1983). Some remarks on Gelb's theory of writing. *General Linguistics*, 23, 198–201.

Swiggers, P. (1984). On the nature of West-Semitic writing systems. *Aula Orientalis*, 2, 149–151.

Trubetzkoy, N. S. (1949) [1939]. *Principes de phonologie*. (Jean Cantineau Trans.). Paris: Klincksieck.

Whitney, W. D. (1889). *Sanskrit grammar*. Cambridge, MA: Harvard University Press.

FOOTNOTES

*To appear in P. Downing, S. Lima, & M. Noonan (Eds.), *Language and Literacy* (Papers from the Milwaukee Symposium on Language and Linguistics).

[1] I will use the neutral term segment in this paper to refer to an element representing a steady state, regardless of assumptions regarding its formal status (e.g., phoneme vs. allophone) in a linguistic model. By segmentation and segmentation ability I mean an exhaustive division of the speech stream into a sequence of segments; ability to recognize some segments does not imply segmentation ability in this, restricted, sense.

[2] It has been argued on a utilitarian basis that incorporation of phonological segments in linguistic analyses is validated by the light shed by such analyses on complex phonological patterns (Anderson, 1974: 6; Halle, 1964: 325). But such arguments lose force in the absence of competing non-segmental analyses of comparable phenomena.

[3] The segments identifiable in an acoustic waveform, to the extent that their boundaries are well-defined, are not isomorphic with the segments of phonological analysis or of alphabetic writing.

[4] See the overview in van der Hulst and Smith (1982).

[5] See Fudge (1987) for an up-to-date review of evidence supporting a hierarchically structured syllable. Cf. Davis (1987).

[6] For discussion of prenasalized consonants, see Anderson (1976), Browman and Goldstein (1986), and Herbert (1986). As Herbert notes, similar considerations apply also to affricates and diphthongs.

[7] The older Japanese children could, in fact, perform some segment manipulation tasks, despite non-exposure to Romaji (Japanese written in the Western alphabet). In her attempt to account for this apparent anomaly, Mann does not consider the possibility that a grid-like presentation of the *kana* syllable signs, in which signs for syllables containing the same vowel are listed in the same column, might provide a passive exposure to segmentation.

[8] This is not to say that failure after appropriate instruction to acquire segmentation ability might not reflect an underlying deficit, as described by Liberman and Shankweiler (1987: 213).

[9] A note on terminology is in order. Early Near Eastern textual material is available in four different segmental scripts: Phoenician, Aramaic, South Arabian, and Ugaritic. These represent texts in four distinct Semitic languages, but the historical relationships among the scripts (for which see Naveh 1982) do not mirror those among the languages. The Aramaic and Phoenician scripts diverged quite early (c. 1000 BC), so it is likely that they had a common ancestor, unfortunately unattested. The South Arabian script, used for inscriptions in now dead languages spoken on the Arabian peninsula, also developed from this common ancestor. The Ugaritic orthography, in contrast, is a segmental cuneiform. Its symbols do not resemble those of the other segmental orthographies, although its citation order was segmentally isomorphic with that of the Phoenician and Aramaic inventories. This suggests a still earlier orthographic tradition, which developed in two distinct ways. Strictly speaking, it is appropriate to use the term *Canaanite* for the script antecedent to the Aramaic and Phoenician scripts, and *West Semitic* for the segmental tradition antecedent both to these scripts and to the Ugaritic cuneiform, but this distinction is not maintained in the literature; Naveh (1982) consistently uses the term *West Semitic* where I would use *Canaanite*. In what follows, I will use the terms *Canaanite* and *West Semitic* for unattested segmental scripts isomorphic (*modulo* segment inventories) to the attested Phoenician, Aramaic, and South Arabian scripts.

[10] Sampson treats the Korean Han-gul as an orthography that codes (distinctive) features; my omission of that category here reflects not disagreement with his analysis, but rather the irrelevance of this category for present purposes. It should be noted that analytic recognition of the featural basis of Han-gul does not imply a claim that Korean speakers make use of this sub-segmental information in decoding the script.

[11] See Browman and Goldstein (1987: 11) for a discussion of the figure-ground metaphor as a structural representation for qualitative differences between consonant and vowel articulation within the model of articulatory phonology.

[12] Non-English fonts in this paper are SuperGreek, SuperHebrew, Sabaean, Ethiopic, and Zakkur (based on a late ninth century BC Aramaic inscription), all copyright by Linguists' Software, and LaserPerfect Phonetique, copyright by Neoscribe International.

[13] The non-linear nature of the Ethiopian script is masked by the subset of symbols presented by Sampson. In all cases, the symbols he shows have right-side vowel marking, preserving linearity. Modification for -ə and -a, however, often shows left-side deviation from the neutral shapes. Note also that this modification may involve removal of material, as in ላ<la>, or displacement of the graph, creating the illusion of left-side deletion, as in ጣ<t'a>. Sampson does not discuss Devanagari, but *conceptually* it is identical to the Ethiopian script; for details, see Whitney (1889: 9-14); cf. Mohanan's (1986: 195) description of the Malayalam orthography as "a syllabary that is 'phonemic'."

[14] I have no information as to whether *ambisyllabic* consonants, consonants which constitute at one and the same time the coda of one syllable and the onset of the following one, occur in languages which are written syllabically.

[15] But (*pace* Firth, 1957: 126; Gelb, 1962) they are not syllabaries either. See below for further discussion.

[16] See Morag (1972) for the origins and structure of these diacritic vocalization systems.

[17] While my definition of alphabet might not accord with popular, non-scientific usage (it is common to speak of the Hebrew or the Arabic alphabet), it is, I think, justified. Popular usage, it should be noted, does not provide any term but alphabet for the writing system of a literate culture, except, perhaps, for "picture writing" and "hieroglyphics" in reference to the Chinese and Egyptian systems.

[18] Mohanan (1986: 197), in a very different theoretical context, offers a compatible definition of an unmarked writing system as one in which each orthographic distinction corresponds to exactly one phonemic distinction in the lexical alphabet; the lexical alphabet is the set of symbols in underlying phonological representation.

[19] Labeling the West Semitic scripts syllabaries may involve a non-standard notion of syllable. If the orthographic form

<q-b-l-t>, representing /qVbbVltV/, is interpreted as a sequence of four syllable signs, then <b> must be taken as representing /.bbV./ and <l> as /.l./, implying a counter-intuitive syllabification /qV.bbV.l.tV/. This is reminiscent of the articulatory syllable of Kozhevnikov and Chistovich (1966: 129ff), which takes as a production unit a vowel and all consonants intervening between it and the preceding vowel (or the beginning of the utterance), differing only in treating /l/ as a discrete syllable.

Hayward and Hassan (1981) describe a curious syllabically linear orthography which was developed for Oromo (Galla), a Cushitic language of Ethiopia, in the 1950's by Shaikh Bakri Sakalo but which never replaced the Amharic-based orthography for that language. The Shaikh's system, unlike the general Ethiopian system, explicitly differentiates single from geminate consonants, writing *babba* <ba-bba>. Mohanan (1986) notes similar syllabifications by Malayalam speakers, both literate and nonliterate.

[20] The relative usage of orthographic types was computed on the basis of population figures and literacy rates in the current *World Almanac*: for example, given the population of Burma (most recent estimate 38,000,000 people) and a reported literacy rate of 66%, 25,080,000 Burmese use a segmentally coded, syllabically linear orthography. The literacy rates in the *Almanac*, especially for developed countries, are suspiciously high; the United States, for example, is reported to have 99% literacy.

[21] For a description of Linear B, a syllabic script representing Mycenean Greek (c. 1500 BC), see Chadwick (1967:75ff). The typological properties of the related script, Linear A, cannot be ascertained, since it has not been determined what language it represents (Packard, 1974; cf. Gordon, 1982: 43ff).

[22] See extensive discussion in Allen (1974: 33ff).

[23] I ignore here Z *zeta*, an affricate /dz/ in early Greek (despite its later Classical value of /zd/ [Allen 1974: 55]). This letter preserves the affricate value of its Semitic source 𐤆 *zain*, a value argued for in Faber (1985, MS). Likewise, use of Semitic 𐤎 *samekh* for Greek Ξ /ks/ (*ksi*) reflects the original affricate value /ts/ of the Phoenician symbol; the directly attested value for cognate sounds in related languages is exclusively /s/, but there is indirect evidence, summarized in Faber (MS), for an earlier affricate value.

[24] The earliest attested West Semitic representation of vowels (the Akkadian syllabically coded system of course allowed representation of vowel sounds, although not in isolation) is in eighth century BC Aramaic texts in which some final vowels and etymologically long vowels are represented ([u]-quality vowels by 𐤅 /w/, [i] quality vowels by 𐤉 /y/, and [a] quality vowels by 𐤀 / or ? 𐤄 /h/). See Bange (1961) for particulars, but with a different interpretation.

[25] Comparable reinterpretations are extremely common in lexical borrowings; borrowed words are regularly restructured in accord with the host phonological system. So, for example, initial *s*C- clusters in English loans into Shona, a Bantu language of Zimbabwe, are regularly nativized with /č/; *spoon* becomes *čipunu* (Chikanza, 1986: 20).

[26] The differing values for *A alpha* and O *o(micron)* in Greek may constitute previously unrecognized evidence for the backing effect of /ʕ/ on an adjacent low vowel in Canaanite; comparable effects are commonly observed in Levantine Arabic dialects, even though in some Maghrebi dialects /a/ tends toward /ɛ/ in such a context. It is unlikely that the value [ɑ] for *ʕayn* reflects the Phoenician change of *á to [ɔ] in stressed, originally open syllables, which Harris (1939: 61) places earlier than the Greek adoption of the Phoenician orthography (on the basis of the letter name *iota*); *a in the Canaanite letter name is in a historically closed syllable and thus would not have been subject to this change. Cohen (1982) suggests that the semantic equation of Canaanite *ʕayn* and Greek *opthalmos*, both 'eye', might have led to the Greek value for O.

[27] The remaining three symbols in Figure 5 B represent sounds not found (or at least not written) in Phoenician. While Naveh does not note this, Φ *phi*, X *khi*, and in their archaic forms (not those forms given in Figure 5) show a striking resemblance to early South Arabian characters 𐩥 and 𐩢, conventionally interpreted as representing /w/ and /ḥ/ (/v/ is a conceivable alternate reading of the former). Φ *phi* presumably represented /pʰ/; Allen (1974: 21) notes that the "classical" value was clearly /pʰ/, but also (p. 23) the possibility of a dialectal pronunciation as /f/ as early as the sixth century BC. X *khi* and Ψ *psi* (more similar in their archaic forms) represented /ks/ and /kʰ/ in the Western scripts from which the Latin script descended and /kʰ/ and /ps/ in the eastern and classical scripts. At least some of the interchange between <w> and <y> symbols in Canaanite, South Arabian, and Greek reflects the Proto-Canaanite shift of *w—>y word initially; the only instances of word initial /w/ in Classical Hebrew and Phoenician are in the morpheme *w-* "and" and in the letter name *waw*, all other instances of initial *w having changed to /y/. Thus, Hebrew *yɛlɛð* "child" is cognate to Arabic *wald*.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 41-73

Phonemic Similarity Effects and Prelexical Phonology*

Georgije Lukatela† and M. T. Turvey††

Ten experiments were conducted on visually presented Serbo-Croatian words and pseudowords, comprising phonemically similar and dissimilar context-target sequences. There were five main results. First, phonemic similarity effects in both lexical decision and naming are independent of graphemic similarity. Second, phonemic similarity need not facilitate lexical decision; the direction of its effect depends on lexicality, target frequency, and type of similarity (specifically, the position of the phoneme that distinguishes context and target). Third, phonemic similarity expedites the naming of words and pseudowords, and to the same degree. Fourth, phonemic similarity is negated in naming but not in lexical decision when the visually presented context and target are stressed differently. Fifth, the phonemic similarity effect occurs even when the context is a masked pseudoword. These results were discussed in terms of a model in which word processing units are activated routinely by phoneme processing units, and compositionally similar word units, when activated, inhibit one another in proportion to each's familiarity. In this model, the phonemic similarity effect in naming is based on the states of phoneme units, whereas the phonemic similarity effect in lexical decision is based on the states of word units. Overall, the results comport with an account in which phonology is computed prelexically and automatically.

Codes for visual lexical access

Enduring questions in the study of word perception concern phonological information and its role. Is phonological information made available routinely in visual lexical access? Is phonological information generated pre-lexically or post-lexically? If phonological information mediates visual access to a word's representation, is the mediation obligatory or optional? There is evidence to suggest that visual word recognition may be mediated by phonological recoding (e.g., Meyer, Schvaneveldt, & Ruddy, 1974). At the same time, strong arguments can be found for the contrary view, namely that visual word recognition can proceed independently of phonological recoding (e.g., Forster & Chambers, 1973). These contrasting observations have led to the development of dual-route models. Roughly, the idea is that the skilled reader of English has an option of using either the visual or phonological route (McCusker, Hillinger, & Bias, 1981). There is a marked tendency, however, to de-emphasize the phonological contribution to word recognition. The deemphasis takes two forms. One promotes visual access as the faster of the two routes and, therefore, the route on which successful reading depends most commonly. The other underscores that most, if not all, of the evidence for phonological processes comes from negative decision data, suggesting that the main contribution of phonological processes is to the perception of letter strings that are pseudowords (Coltheart, Davelaar, Jonasson, & Besner, 1977).

In the dual-route model the visual and phonological routes operate in parallel, with the phonological code derived on the basis of "grapheme-to-phoneme correspondences" or GPCs. The idea is that GPCs operate by breaking down a word into its constituent graphemes and assigning each one its appropriate phonemic code. It has been argued

---

This research was supported in part by National Institute of Child Health and Human Development Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively. The second author held a James McKeen Cattell Fellowship during the preparation of this manuscript.

that if prelexical phonological recodings are produced by using GPC rules, then such recodings can only occur for words that conform to these rules, that is, so-called "regular words" (Coltheart, Besner, Jonasson, & Davelaar, 1979). The lexical entries of words that are exceptions to the rules can be accessed only via the visual route. Consequently, whether or not the lexical decision task involves the use of GPCs might be evaluated by determining whether performance is better with regular words than with exception words. Unfortunately, regularity effects in English word processing prove to be elusive when measured in lexical decision, with the result that the evaluation of the GPC process in terms of regularity effects has been inconclusive (see Humphreys & Evett, 1985). More convincing evidence for GPCs is to be found in studies with Serbo-Croatian.

Visual word recognition in Serbo-Croatian

The script system of Serbo-Croatian contrasts in two important ways with that of English. First, it occurs in two forms, namely, Roman and Cyrillic alphabets. These two alphabets share a number of letters (in upper case: A, E, O, J, K, M, T, H, P, C, B), some of which (H, P, C, B) designate one phoneme in one alphabet and another phoneme in the other alphabet (e.g., Lukatela & Turvey, 1980). Second, individual letters within an alphabet map uniquely to individual phonemes and individual phonemes map uniquely to individual letters; the orthography is shallow (see Frost, Bentin, & Katz, 1987; Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela & Turvey, 1980). Previous research has suggested that a prelexically derived phonology plays a major role in word recognition and word naming in Serbo-Croatian (e.g., Feldman & Turvey, 1983; Frost et al., 1987; Lukatela et al., 1980). The empirical support has come primarily from the phonological ambiguity effect.

There are written words in Serbo-Croatian that comprise only shared letters. Some of these contain phonemically ambiguous letters. Consider the letter string BETAP. Read strictly through the letter-to-sound correspondences of the Cyrillic alphabet, this letter string is pronounced /vetar/ and is a high frequency noun meaning "wind." Read strictly through the letter-to-sound correspondences of the Roman alphabet, BETAP is pronounced /betap/, a pseudoword. Read with a mixture of the two sets of correspondences, Cyrillic and Roman, leads to the pronunciations /vetap/ and /betar/, which are also pseudowords. The word meaning "wind" is transcribed in the Roman alphabet as VETAR. This letter string supports only a single reading, /vetar/. No other readings are possible. VETAR, unlike its Cyrillic mate BETAP, is phonologically unambiguous.

In the lexical decision, the acceptance latencies for BETAP, and for words like it, are considerably longer than the acceptance latencies for VETAR, and for words like it, even though BETAP and VETAR are equal in frequency, syllabic structure, number of letters, and meaning. This contrast defines what is referred to as the phonological ambiguity effect (PAE). It is argued that accepting BETAP as a word is made more difficult by the fact that it activates more phoneme units and, therefore, activates partially more competing word units, than VETAR (Lukatela, Turvey, Feldman, Carello, & Katz, 1989). Importantly, PAE satisfies Coltheart et al.'s (1977) and Van Orden's (1987) criterial requirement for the demonstration of phonological mediation in lexical access: Phonological effects are demonstrated in positive lexical decisions. The effect of phonological ambiguity on "yes" responses means that phonology cannot be considered as a representation that is delayed relative to visual lexical access. Let us turn to the major supplementary features of PAE. In our view, they provide a stubborn set of facts favoring the automatic involvement of prelexical phonology in lexical access.

(1) *PAE occurs for pseudowords.* Rejection latencies are slowed for the phonologically ambiguous pseudoword BEMAP relative to its phonologically unambiguous pseudoword mate VEMAR. Arguably, BEMAP is coded into more phonological forms than VEMAR and BEMAP's rejection latencies are, therefore, slowed relative to those for VEMAR. If pseudowords do not have lexical entries, as commonly argued, then their phonological representations must be derived prelexically through letter-phoneme correspondences, or "postlexically" by analogy (Glushko, 1979; Kay & Marcell, 1981). Because PAE occurs for both words and pseudowords, and because PAE is due to phonological factors, then it seems likely that phonological information is activated automatically given a letter string.

(2) *PAE is generally larger for words than pseudowords.* With one exception (Lukatela, Carello, Savić, & Turvey, 1986, using lateralized displays), PAE has proven to be more pronounced in word than pseudoword lexical decision times. On the basis of English language research the argument has been made that phonological coding must be optional and rare in ordinary word

recognition because phonological factors influence pseudoword processing more so than word processing (Coltheart et al., 1977, Henderson, 1982; McCusker et al., 1981). Applying this logic to the Serbo-Croatian results, phonological coding must be mandatory and routine in ordinary word recognition because phonological factors influence word processing more so than pseudoword processing (e.g., Feldman & Turvey, 1983; Lukatela et al., 1980; Lukatela, Feldman et al., 1989).

(3) *PAE occurs in both lexical decision and naming.* As noted, lexical decision and naming appear to be sensitive differentially to pre- and post-lexical influences. It is argued that effects obtained with lexical decision but not naming are postlexical effects, and those obtained with both tasks are prelexical (Balota & Chumbley, 1985; Seidenberg, et al., 1984; West & Stanovich, 1982). PAE is found repeatedly in both lexical decision and naming tasks (e.g., Lukatela, Feldman et al., 1989; Lukatela, Turvey, et al., 1989). According to the preceding criteria, PAE is prelexical.

(4) *PAE can be reduced by an immediately preceding context that specifies the alphabet of BETAP-type words.* The difference between lexical decision latencies to BETAP and VETAR is minimized or eliminated when BETAP is preceded by a Cyrillic letter string (that is, a context written with unique Cyrillic letters specifying, thereby, the same alphabet as that in which BETAP is a word) and VETAR is preceded by a Roman letter string (Lukatela, Turvey, et al., 1989). Importantly, PAE is reduced equivalently by a nonword, consonant string context with no orthographic resemblance to BETAP that specifies the alphabet in which BETAP is a word, and a word context that does the same thing (Lukatela, Turvey, et al., 1989). Apparently, some process at the level of letter units can reduce the number of phonemic representations of BETAP from four to one, thereby rendering it similar to VETAR. The implication is that the origin of PAE is prelexical.

(5) *PAE is not eliminated by experimental conditions that encourage suppressing phonological coding.* In one experiment subjects were encouraged to process all items as written in Roman (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978, Experiment 1). Unique Cyrillic letters never appeared. The phonologically ambiguous words were words in Roman and nonwords in Cyrillic; they comprised only 9% of the total stimuli. Another 9% of the stimuli were phonologically ambiguous pseudowords, making only a total of 18% that could be read alternatively as Cyrillic. In sum, the experiments of Lukatela et al. (1978) encouraged using only the Roman letter-phoneme correspondences, or using a visual word-specific code. PAE was, however, pronounced. In another experiment, the phonologically ambiguous letter strings that were words, were words in their Cyrillic reading and all the subjects were people who had learned Cyrillic as their first alphabet (Feldman & Turvey, 1983). The optimal strategy would have been to suppress the (Roman) nonword reading in favor solely of the (Cyrillic) word reading, or simply to use a visual word-specific code. Again, PAE was pronounced. In English language experiments that involve phonologically ambiguous stimuli (homophones, pseudohomophones), high percentages are recommended to avoid biasing the subject to a phonological coding strategy (Davelaar, Coltheart, Besner, & Jonasson, 1978; Hawkins, Reicher, Rogers, & Peterson, 1976). The idea is that when the experimental conditions are unfavorable to the use of phonological coding—it results in slowed and inaccurate responses—then phonological coding is abandoned in favor of strictly visual access. Serbo-Croatian experiments have been conducted in which 50% of all stimuli were phonologically ambiguous, half of the time in Cyrillic and half of the time in Roman (Lukatela, Feldman, et al., 1989). Using visual access would have avoided the slowing down and errors induced by the phonological code. Once again, PAE was in evidence. Clearly, PAE is indifferent to wide variations in experimental conditions (e.g., 12.5% ambiguous stimuli in Feldman et al. [1983] to 50% ambiguous stimuli in Lukatela, Feldman, et al. [1989]). The implication is that PAE is due to an automatic process, not an optional or strategic process, and that the automatic process involves phonology.

(6) *PAE is not the consequence of phonologically ambiguous words being associated with intrinsically slower responses due to lexical entries with higher activation thresholds.* It has been hypothesized that visual access usually provides enough information for lexical decisions to be made, but sometimes the accessing is sufficiently slow (e.g., the word is infrequent, the spelling is irregular, the subject is tardy) to allow lexically derived phonology a role in the decision process (Seidenberg, 1985). To reiterate, the difference between lexical decision latencies to BETAP and VETAR is minimized or eliminated when BETAP is preceded by a Cyrillic letter string (that is, a context in the same alphabet as that in which BETAP is a word) and VETAR is preceded by a

Roman letter string (Lukatela, Turvey, et al., 1989). The implication is that there is nothing especially different about the lexical entries of BETAP and VETAR, or their spellings, or the kinds of information they make available postlexically. PAE seems to depend on prelexical rather than postlexical processes.

(7) *PAE is independent of familiarity/frequency.* The inflected nouns of Serbo-Croatian are most frequent and most prominent grammatically in the nominative singular form. A common finding is that nominative singulars are responded to faster than oblique forms (e.g., Lukatela, Carello, & Turvey, 1987). For example, VENA ("vein" in nominative singular) is responded to much faster than VENI (same word in dative or locative singular). However, if VENA and VENI are written in Cyrillic, that is, BEHA and BEHИ, respectively, then the less frequent and less grammatically prominent dative/locative form BEHИ is responded to faster. The latency difference between BEHA and VENA is large and defines PAE; the latencies to BEHИ and VENA do not differ (Feldman, et al., 1983). Further, if latency to the unambiguous form of a word is taken as a measure of experiential familiarity with the word, then PAE in lexical decision does not vary in magnitude with experiential familiarity (Feldman & Turvey, 1983). The implication is that PAE is not governed by biases (e.g., activation thresholds, grammatical significance) at the level of lexical entries.

(8) *PAE depends on the number of ambiguous letters.* Effects dependent on the number of letters or syllables in a word have been used as indices of phonological recoding (Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Green & Shallice, 1976). Letter length effects occur regularly in naming but rarely, if ever, in lexical decision (Henderson, 1982). The lack of such effects in lexical decision is used to argue that phonological coding cannot take place. KOTBA has one ambiguous letter (B), BETAP has two ambiguous letters (B and P), and CABAHA has three (C, B, and H). In each example the remaining unambiguous letters are shared letters. In the lexical decision task with such stimuli, PAE is larger the greater the number of ambiguous letters (Feldman et al., 1983; Feldman & Turvey, 1983). Conversely, other research shows that the effect in lexical decision is smaller the greater the number of unique letters (Lukatela, Feldman, et al., 1989). The implication is that PAE is based on an analytic process that heeds the phonological aspects of the printed word as specified in the details of its orthographic structure (Turvey, Feldman, & Lukatela, 1984).

(9) *PAE can appear as false negative responses.* Rejecting a BETAP-type word as a word occurred in the Serbo-Croatian experiments on a high proportion of the times (in the range of 19% to 26%) such a word was presented, though in those experiments average false negatives on phonologically unambiguous words were in the range of 1% to 4%. High average of the false negatives responses on phonologically ambiguous words suggests a mental process in which the representation activating the lexicon does not correspond to a word. This is understandable if PAE is due to the facts that (a) BETAP can give rise to a number of phonological representations that do not correspond to a word, and (b) the lexicon is accessed through phonological representations.

(10) *PAE can appear as false positive responses.* There are two significant types of phonologically ambiguous pseudowords. One type, like BEMAP, comprises letter strings that are (a) composed of shared letters, one or more of which are ambiguous, (b) unreadable as a word in either Roman or Cyrillic, and (c) unreadable as a word in a mixture of both alphabets. The other type satisfies (a) and (b) but not (c). An example of the second type is the pseudoword HAPEB. If the letter-phoneme correspondences of both alphabets are applied to HAPEB, then one of the resulting eight phonological descriptions corresponds to a word, namely, /napev/ meaning "tune." It comes about by assigning the phoneme /n/ to H by the Cyrillic alphabet, the phoneme /p/ to P by the Roman alphabet, and the phoneme /v/ to B by the Cyrillic alphabet. Compare HAPEB to a pseudoword of the first type, such as BEMAP. BEMAP has all but one letter in common with a real word (BETAP), and all but one phoneme in common with this real word. HAPEB similarly has all but one letter in common with a real word (HAПEB). However, HAPEB has all phonemes in common with a real word, if both alphabets apply.

If lexical access is visual, then false positives to BEMAP and HAPEB should be few and equal. In contrast, if lexical access is based on phonemic descriptions computed prelexically, and if the computation is analytic, assigning as many phonemes per letter as permitted, then false positives should be greater for HAPEB. Experimentation shows that BEMAP and its unambiguous pseudoword control VEMAR generate the same small number of false positives (about 2-3%). When preceded by a neutral context word, HAPEB generated 31% false positives;

when preceded by an associate of /napev/, namely, MELODIJA ("melody"), HAPEB generated 55% false positives (Lukatela, Turvey, et al., 1989). The same pattern of results held for naming (Lukatela, Feldman, et al., 1989; Lukatela, Turvey, et al., 1989). These results are understandable if PAE with pseudowords is due to the facts that (a) a phonologically ambiguous pseudoword can give rise automatically to all of the phonological representations that its letter structure permits, (b) for some phonologically ambiguous pseudowords one of the automatically generated phonological representations corresponds to a word, and (c) the lexicon is accessed through phonological representations.

Let us put the preceding survey of PAE features in perspective. Various experimental criteria have been advanced by investigators of English word processing for deciding whether or not phonological coding mediates the accessing of word information. The criteria were incorporated in the preceding survey. These experimental criteria have not been satisfied in English language studies, with the one recent and important exception of Van Orden's research (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988). In contrast, most of the criteria have been satisfied in Serbo-Croatian language studies.

Facilitating word recognition by phonemic similarity

One strategy for addressing questions of prelexical phonology is to examine how the visual processing of a given word (the target word) is influenced by the degree of shared phonemic and graphemic similarity between it and a preceding word (the context word). In the original application of this strategy, using lexical decision latency as the measure of word processing, it was observed that TRIBE following BRIBE was responded to faster than TRIBE following FENCE, and that BREAK following FREAK was responded to more slowly than BREAK following COUCH (Meyer, Schvaneveldt, & Ruddy,1974). Meyer et al. (1974) interpreted these results to mean that phonological information is automatically accessed in visual word recognition and, specifically, that people tend to use the same letter-phoneme conversion rules on the target as were used on the context. Meyer et al. (1974) argued for an encoding bias hypothesis: If a phoneme code mediates lexical access, then lexical decision should be more rapid when the letter-to-phoneme mappings are similar for context and target (BRIBE-TRIBE) and more slow when the letter-to-phoneme mappings are dissimilar for context and target (FREAK-BREAK).

Subsequent research has raised doubts about the encoding bias hypothesis. First, it appears that the TRIBE-BRIBE effect and the BREAK-FREAK effect result from different mechanisms. When the stimuli requiring a "no" response are orthographically illegal strings rather than pseudowords (orthographically legal, pronounceable nonwords), the inhibition of BREAK by FREAK disappears (Hanson & Fowler, 1987; Shulman, Hornak, & Sanders, 1978). The BREAK-FREAK effect also disappeared when a delay of the order of 250 ms intervened between context and target (Hillinger, 1980). In contrast, the phonemic similarity effect—the facilitation of TRIBE by BRIBE—persisted over these manipulations of context type and time. Second, the phonemic similarity effect endured when the context was presented auditorily, and it was of the same magnitude whether the context was graphemically dissimilar (EIGHT-MATE) or graphemically similar (LATE-MATE) (Hillinger, 1980).

Recently, Martin and Jensen (1988) conducted a series of five visual lexical decision experiments in which they used word targets preceded by graphemically similar or graphemically dissimilar primes that rhymed with the target. In none of these experiments was the phonemic similarity effect observed. The failure to replicate the most important of Hillinger's (1980) results led Martin and Jensen (1988) to suggest that either (a) the internal lexicon is not organized such that spreading activation occurs on the basis of phonemic similarity, or (b) visual lexical decisions are made without phonological mediation. Other kinds of experiments have questioned the automatic activation of phonological information on which the encoding bias hypothesis is based. The observation of longer latencies in the regular lexical decision task to homophonic words such as YOKE (Rubinstein, Lewis, & Rubinstein, 1971) fails to hold when such words occur rarely in the experiment (Davelaar et al., 1978). More robust is the pseudoword homophonic effect in which TRATE type items yield slower rejection latencies. It has been argued, however, that because the latter is a phonological effect in slower (nonword) responses, it is unlikely to occur in faster (word) responses. That is, word processing may occur ordinarily without a prelexical phonological code (Coltheart et al., 1979; Henderson, 1982; McCusker, Hillinger, & Bias, 1981). New results, however, are more favorable to the view that phonology mediates lexical access in English word recognition (Perfetti, Bell, & Delaney, 1988; Van Orden, 1987; Van Orden et al., 1988).

Investigating the phonemic similarity effect in Serbo-Croatian

In the present article we investigate the phonemic similarity effect and pursue the main points of Meyer et al.'s (1974) original thesis, namely, that a prelexical phonology is involved in fluent word processing, and that it can be biased by the phonological features of preceding contexts. The present article is distinguished from previous research on the phonemic similarity effect by its use of both lexical decision and naming tasks, and by its use of Serbo-Croatian—a language, as noted above, that differs from English in terms of the relative complexity of computing phonology from script.

The influence of contextual information on visual word processing tasks has been found to depend not only on the kind of context but also on the nature of the task. The lexical decision and naming tasks have emerged as necessary converging operations to be considered in any evaluation of context effects, because they appear to be differentially sensitive to pre- and postlexical influences (Carello, Lukatela, & Turvey, 1988; Forster, 1979; Seidenberg, Waters, Sanders, & Langer, 1984; Stanovich & West, 1983; West & Stanovich, 1982, 1986). The argument is that effects obtained with both tasks are prelexical in origin, whereas those obtained only in lexical decision are postlexical in origin, where "prelexical" refers to processes of activating word information and "postlexical" refers to processes that use this information (Balota & Chumbley, 1985; Seidenberg et al., 1984; West & Stanovich, 1982). The fundamental difference between the two tasks can be expressed in signal detection terminology (Seidenberg et al., 1984): Lexical decision involves sensitivity and bias; naming involves only sensitivity. The point is that in naming there is no criterion to be referred to because there is no decision to be made.

With respect to the relevance of the phonemic similarity effect to issues of phonological coding, the most significant finding would be that it was independent of graphemic similarity: EIGHT primes MATE. This result of Hillinger (1980), though seriously questioned by Martin and Jensen (1988), would suggest that the phonemic similarity effect results strictly from the use in target processing of phonological information activated by the context. Hillinger's (1980) demonstration that EIGHT primed MATE, compared EIGHT-MATE (phonemically similar, graphemically dissimilar) with the pairs LATE-MATE (phonemically similar, graphemically similar) and VEIL-MATE (phonemically dissimilar, graphemically dissimilar). With respect to lexical decision times, the two rhyming pairs did not differ from each other, and both differed from the control. A later test of independence from graphemic similarity was conducted by Evett and Humphreys (1981) using a masking procedure and percent correct identification. They compared the difference between FILE-TILE and its control with the difference between COUCH-TOUCH and its control and found them to be the same, concluding thereby that the increased recognition of TILE following FILE could be accounted for strictly in terms of graphemic factors. In other masking experiments (Humphreys, Evett, & Taylor, 1982), and in lexical decision experiments (Hanson & Fowler, 1987), it has been the case, however, that orthographically similar rhyming pairs show greater facilitation than orthographically dissimilar nonrhyming pairs.

Where different word pairs are contrasted, the outcome can be contaminated by frequency, semantic, and other uncontrolled differences. Automatic postlexical checks for associativeness, syntactic agreement, and semantic coherence could differ between pairs like FILE-TILE and COUCH-TOUCH. In the Hillinger (1980) study, EIGHT-MATE and LATE-MATE are different word pairs and are, therefore, obviously different in other ways than graphemically dissimilar vs. graphemically similar. Consider as further examples of Hillinger's stimuli, HEARD-BIRD and THIRD-BIRD, FRUIT-ROOT and LOOT-ROOT, WEIRD-BEARD and GEARED-BEARD, SEWS-HOSE and ROSE-HOSE. These contrasts on the dimension of graphemic similarity include also contrasts in syntactic coherence, associativeness, frequencies of the context items (e.g., LOOT = 3/million, FRUIT = 35/million, Kučera & Francis [1967]), and the word types that are paired (e.g., verb-noun vs. number-noun, adjective-noun vs. verb-noun, and so on). A distinct advantage of Serbo-Croatian is that it provides a way of evaluating the independence of the phonemic similarity effect from graphemic similarity with the same word pairs. This is achieved by writing the context word and the target word in the same alphabet, for example, PUTIĆ ("lane" written in Roman) - PUŽIĆ ("snail"

written in Roman), or in different alphabets, ЛУТИЋ ("lane" written in Cyrillic) - PUŽIĆ. In sum, Serbo-Croatian permits a pure test of the independence of the phonemic similarity effect from graphemic similarity. If the phonemic similarity effect passes the test, then the claim that phonological information underwrites this phenomenon of visual word processing would be secure. The focus of inquiry can then turn to the issue of where and when in word processing such information is made available.

EXPERIMENT 1

The first experiment was an attempt to replicate for Serbo-Croatian Hillinger's (1980) observation that the successive visual presentation of two phonemically similar words affects the lexical decision to the second word and does so independently of graphemic similarity. On each trial lexical decision was made on a target preceded by a context that was phonemically similar (differing from the target only in the middle consonant) or dissimilar. Also manipulated was the graphemic similarity between context and target. This was done by presenting the context in either the same alphabet as the target or in the other alphabet. Importantly, this way of manipulating graphemic similarity preserves the same phonological relation over graphemically similar sequences, for example, PUTIĆ-PUŽIĆ (Type Ib, see Table 1) and graphemically dissimilar sequences, for example, ЛУТИЋ-PUŽIĆ (Type IIb), where PUTIĆ and ЛУТИЋ are the same context word written in Roman and Cyrillic, respectively. To show the phonemic similarity effect under conditions of graphemic dissimilarity is to provide evidence that phonological information is automatically accessed; target lexical decision is affected by the the context word's phonology, rather than by the context word's orthography or form. In the present study the phonemic similarity effect will be tested by exploiting four stimulus types that are illustrated in Table 1. Contexts and targets were: I. phonemically and graphemically similar, written in the same alphabet; II. phonemically similar and graphemically dissimilar, written in different alphabets (half of targets were Roman and half were Cyrillic); III. phonemically and graphemically dissimilar, written in the same alphabet; and IV. phonemically and graphemically dissimilar, written in different alphabets (half Roman and half Cyrillic). Phonemically dissimilar pairs (Type III and Type IV) provide a baseline for the corresponding phonemically similar pairs (Type I and Type II, respectively).

Table 1a. *Examples of phonemically similar words in the same alphabet (Type I pairs) that share all letters/phonemes except for the initial (a), middle (b), or final letter/phoneme (c).*

Type I	Examples	
Ia	KUVAR-ČUVAR (cook-guardian)	ЖИВАЦ-ЛИВАЦ (nerve-smelter)
Ib	PUTIC-PUŽIC (lane-snail)	РАКУН-РАЧУН (racoon-bill)
Ic	GUSAR-GUSAN (pirate-gander)	ПУТИЋ- ПУЖИЋ (lane-snail)

Table 1b. *Examples of phonemically similar words in different alphabets (Type II pairs) that differ either in the initial (a), middle (b), or final phoneme (c), and which share no more than one letter.*

Type II	Examples	
IIa	КУВАР-ČUVAR (cook-guardian)	ŽIVAC-ЛИВАЦ (nerve-smelter)
IIb	ПУТИЋ-PUŽIC (lane-snail)	RAKUN-РАЧУН (racoon-bill)
IIc	ГУСАР-GUSAN (pirate-gander)	PASUS-ПАСУЉ (paragraph-beans)

Table 1c. *Examples of phonemically dissimilar words in the same alphabet (Type III pairs) and phonemically dissimilar words in different alphabets (Type IV) with no more than one shared letter/phoneme.*

Type III	Type IV
PASUS-ČUVAR (paragraph-gardian)	ПАСУС-ČUVAR (paragraph-gardian)
РАКУН-ЛИВАЦ (racoon-smelter)	RAKUN-ЛИВАЦ (racoon-smelter)
GUSAR-PUŽIĆ (pirate-snail)	ГУСАР-PUŽIĆ (pirate-snail)
КУВАР-РАЧУН (cook-bill)	KUVAR-РАЧУН (cook-bill)
PUTIĆ-GUSAN (lane-gander)	ПУТИЋ-GUSAN (lane-gander)
ЖИВАЦ-ПАСУЉ (nerve-beans)	ŽIVAC-ПАСУЉ (nerve-beans)

Four other features require highlighting. First, English language experiments have used word sets dominated by monosyllabic words; the present experiment uses two and three syllable words primarily because of the rarity of monosyllabic words in Serbo-Croatian. Second, in English experiments phonemic similarity between pseudowords cannot always be defined unequivocally because the pronunciation of pseudowords may be ambiguous. The present experiment uses word-pseudoword pairs for which the phonemic similarity relation is unambiguous. This feature follows from the facts that the Serbo-Croatian writing system (both Roman and Cyrillic) is phonemically precise and that all letters in a word are pronounced. Thus, a word context that differs from a pseudoword target in one letter is phonemically identical to the target with respect to all letter positions but one. Third, context-target pairs that are graphemically similar but phonemically dissimilar (e.g., FREAK-BREAK) can not appear in Serbo-Croatian experiments because the phonetic precision of the Serbo-Croatian orthography does not permit grapheme-phoneme dissociations of the kind represented by FREAK-BREAK. Fourth, phonemic similarity can be of different types depending on the position of the unshared phoneme or phonemes. In Experiment 1, context and target in a phonemically similar pair differ in the middle consonant. They were phonemically similar but do not rhyme according to the standard definition.

Method

Subjects

Seventy-six high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials

Seventy-two word-word pairs were selected that were phonemically and, if written in the same alphabet, graphemically similar. To be considered phonemically similar, they could differ by only one letter. Word pairs were chosen so that this different letter was a consonant in the middle or approximately in the middle. Contexts and targets were chosen to be as similar as possible with respect to accent position (i.e., syllable) and type (i.e., short rising, short falling, long rising, or long falling) and without any associative or semantic relationship. All words were nouns (masculine and feminine), all in the nominative singular grammatical form. They comprised two or three syllables and four, five or six letters, with the majority being two-syllable, five-letter words. Seventy-two pseudoword targets were derived from the word targets by changing the one consonant that was different from the context (so that pseudowords can also be considered to be derived from the contexts).

Examples of context-target pairs in Experiment 1 are given in Table 2. Context words are associated with the type numbers which are explained by Table 1. Each type occurred with both word and pseudoword targets.

Design

A given context was seen by an individual subject twice, once preceding the word target and once preceding the derived pseudoword target. When the word was Roman, the corresponding pseudoword was Cyrillic and vice versa. A given subject never encountered a target word or target pseudoword in any of the pairs more than once but every subject saw every type of pair (I, II, III, and IV), and every target appeared in every type of pair. These constraints were met with four counterbalancing groups. In total, each subject saw 144 experimental stimulus pairs ordered randomly. The experimental sequence was preceded by 32 practice pairs of the same type (although the pairs were less successful in satisfying the constraints on stimulus selection).

Procedure

A subject was seated before the CRT of an Apple IIe computer in a dimly lit room. A fixation point was centered on the screen. On each trial, the subject heard a brief warning signal after which a word appeared for 500 ms above the fixation point. After a 100 ms interstimulus interval, another letter string appeared below the fixation point for 500 ms. Subjects were instructed to decide as rapidly as possible whether or not the second string was a word. To ensure that subjects were reading the contexts, they were asked occasionally by a computer message to report orally both stimuli after the lexical decision had been made (on average, approximately every 15 trials). Decisions were indicated by depressing with both thumbs a telegraph key for a "No" response or by depressing another key, slightly further away, with both forefingers for a "Yes" response. Latencies were measured from the onset of the target. If the response latency was longer than 1500 ms, a message appeared on the screen requesting that the subject respond more quickly, but long decision times were included in the analysis. The intertrial interval was 2500 ms.

Table 2. *Examples of context-target pairs used in Experiments 1 and 5 in which members differ in a letter/phoneme near the middle.*

Target		Context word			
Word	**Pseudoword**	**Type Ib***	**Type III**	**Type IIb**	**Type IV**
PUŽIC	PUDIĆ	PUTIĆ	RAKUN	ПУТИЋ	РАКУН
RAČUN	RAŽUN	RAKUN	PUTIĆ	РАКУН	ПУТИЋ
ЛИКЕР	ЛИГЕР	ЛИДЕР	БУКАЧ	LIDER	BUKAČ
БУХАЧ	БУЉАЧ	БУКАЧ	ЛИДЕР	BUKAČ	LIDER
BUTINA	BUŠINA	BULJINA	RIBANJE	БУЉИНА	РИБАЊЕ
RITANJE	RIFANJE	RIBANJE	BULJINA	РИБАЊЕ	БУЉИНА
СУШЕЊЕ	СУЉЕЊЕ	СУЂЕЊЕ	ЖИЧИЦА	SUĐENJE	ŽIČICA
ЖИЛИЦА	ЖИФИЦА	ЖИЧИЦА	СУЂЕЊЕ	ŽIČICA	SUĐENJE

*Refers to the type of context-target pair identified in Table 1 that is produced by pairing the given context word with the target.

The experimental sequence was preceded by a practice sequence of eight different stimulus pairs of the same four types. After one quarter of the trials, there was a brief rest. The next quarter was again preceded by eight practice fillers, as were the third quarter, and the fourth. Within each quarter, context-target types were ordered randomly.

Results and Discussion

Minimum and maximum acceptable latencies were set at 250 ms and 2000 ms, respectively. Responses outside of this range were counted as errors. The foregoing conventions applied throughout the series of experiments. Table 3 shows the word and pseudoword mean reaction times, standard deviations, and errors. An analysis of variance was conducted with the factors of phonemic similarity (similar/dissimilar), alphabet (same/ other), and lexicality (word/pseudoword). The main effect of phonemic similarity was significant, $min\ F'(1,190) = 4.71$, $p < .05$, but phonemically similar contexts hindered rather than aided lexical decision: similar = 836 ms vs. dissimilar = 823 ms. The analysis also revealed a significant effect of lexicality (words = 803 ms vs. pseudowords = 857 ms), $min\ F'(1, 207) = 32.44$, $p < .001$, but neither the main effect of alphabet (same = 827 ms vs. other = 833 ms), $min\ F'(1, 217) = 1.48$, $p > .05$, nor its interaction with phonemic similarity, $min\ F'(1, 210) = 2.04$, $p > .05$, were significant. Although lexical decision on words appeared to exhibit a larger negative phonemic similarity effect than lexical decision on pseudowords, there was no significant similarity by lexicality interaction, $min\ F'(1, 210) = 1.27$, $p > .05$. For the analysis of variance conducted on the error data only lexicality proved significant, $min\ F'(1, 205) = 15.21$, $p < .001$.

Table 3. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 1.*

Target	**Alphabet**	**Phonemic relation**			
		Similar		**Dissimilar**	
		L	**ER**	**L**	**ER**
Word:					
	same	807[a]	5.26	792	5.48
		97[b]	6.28	105	6.70
		70[c]	6.96	73	7.25
	other	820	5.48	793	5.12
		104	6.32	102	6.20
		64	7.25	59	7.85
Pseudoword:					
	same	851	2.49	854	3.07
		111	3.56	113	4.57
		67	4.33	69	5.21
	other	865	2.85	853	2.34
		106	3.90	100	3.30
		62	4.41	62	3.63

[a] Mean

[b] standard deviation by subjects

[c] standard deviation by items

Obviously, phonemic similarity affected lexical decision on words independently of the alphabetic relation between context and target (e.g., ЛУТИЋ-PUŽIĆ compared to PUTIĆ-PUŽIĆ), demonstrating that the effect was governed by phonemic similarity and not by orthographic or form similarity. However, in contrast to the positive effect observed by Hillinger (1980), the phonemic similarity effect of the present experiment was negative. A major difference between the present experiment and Hillinger's experiments is that his stimuli differed in the initial consonant or consonants (e.g., LOAD-TOAD, THIEF-GRIEF) whereas those of the present experiment differed in the middle consonant.

Analysis of the lexical decision task suggests a number of variables likely to affect the magnitude and direction of the phonemic similarity effect. In approximate terms, a word stimulus presented for lexical decision does the following: It activates feature, letter, and phoneme processing units; it activates, via letter and/or phoneme units, all word units that share components (letters/phonemes) of its structure, with the degree of activation varying with the degree of overlap; it sets off inhibitory activity among activated, compositionally similar word units that trims the number of units competing for recognition (McClelland & Rumelhart, 1981); and it engages a signal-detection capability that evaluates the activity levels of word units against a criterion. When a target word follows a phonemically similar or dissimilar context word, the processes just identified repeat and modulate those that occurred recently before, as well as being modulated by them.

We elaborate on three points in the preceding. First, the activations of processing units by the different parts of a word stimulus need not be in synchrony, for example, initial letters may induce these activities sooner than middle and terminal letters. Relatedly, the letters in a word need not be uniformly informative for identification purposes (Merikle & Glick, 1976), and need not contribute linearly to the phonemic similarity effect (Evett & Humphreys, 1981). Second, increases and decreases in the activity levels of partially activated word units, and the inhibition levels they incur, need not be constant in magnitude and rate but variable, depending on a unit's frequency, number of components, and so on (McClelland & Rumelhart, 1981). Third, the evaluation process need not use a constant criterion of acceptance but one that varies with the ratio of the highest activation level to the average activation level, and with the discriminability of current and recent stimuli (Balota & Chumbley, 1985; Norris, 1986). We ought to expect the magnitude and direction of the phonemic similarity effect, therefore, to be complexly determined.

Three readily manipulated variables of likely significance emerge from the foregoing considerations: (a) the type of phonemic similarity, that is, the letter positions in which the context and target overlap phonemically, (b) the frequency of the target word, and (c) the frequency of the context word. In the second experiment we make a direct comparison between the two different similarity types, initial different and middle different, in order to evaluate the contribution of the position of the unshared phoneme to the direction of the phonemic similarity effect. In the third experiment we repeat the initial different vs. middle different position manipulation, and add a manipulation of target word frequency. In the fourth experiment, the phonemic similarity type comprising words that differ in final consonants is examined, and the frequencies of both context words and target words are manipulated. Each of the experiments employs the lexical decision task.

EXPERIMENT 2

This experiment compares the effects of phonemic similarity for stimuli that differ either in the initial consonant or in the middle consonant, and that differ graphemically (one word in each context-target pair is written in Cyrillic, the other in Roman). There is reason to believe (e.g., Evett & Humphreys, 1981) that the letter positions in a word or pseudoword may not be of equal status in priming situations. We elaborate on this notion below.

For simplicity, consider three-letter stimuli. Four kinds of context-target word sequences can be identified: ABC-ABC (identity); XBC-ABC (initial different); AXC-ABC (middle different); and ABX-ABC (final different). In each situation, the context activates a number of word units. The word unit corresponding to the context is activated fully; other word units are activated partially. These partially activated units have some of the same letters and/or phonemes in the same positions as the context word. Included among them, therefore, will be the unit corresponding to the target. By all accounts, the identity situation produces the largest priming effect, and an effect that is positive. Presumably, this is because the context ABC fully activates the word unit that is used in the lexical decision on the target ABC. In the three different situations (initial, middle, and final) the context activates partially the target unit, and the resulting priming effect is less than maximal, and not necessarily positive (e.g., Experiment 1). We propose the following hypothesis: The activity level of a partially activated word processing unit depends on which parts of the unit are activated. More specifically, we propose that the automatic activation of the word processing unit corresponding to ABC is least during the processing of XBC, most during the processing of ABX, and intermediate during the processing of AXC. (From the point of view of information theory, activation of a word unit's initial phoneme relative to the other phonemes resolves most of the uncertainty concerning the question of which word was presented.) Simply speaking, if the phonemic similarity effect depends on how the target's representation is prepared by the context, then the effect should differ across initial different, middle different, and final different context-target sequences.

The present experiment uses 5 letter stimuli comprising consonants (C) and vowels (V) in the pattern CVCVC. According to the hypothesis just advanced, we should expect a difference in phonemic similarity effects between context-target sequences that share all but the initial C, and context-target sequences that share all but the middle C. In addition, the present experiment examines the potential side effect of stimulus presentation order. Lexical decision on the target word of a given context-target pair could be influenced by the preceding stimulus pair. Although such an outcome was considered unlikely, prudence suggested that its significance be evaluated.

Method

Subjects

Forty-four high school seniors from the Fourth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four groups according to his or her appearance at the laboratory. Subjects of the first two groups participated in "Version A" of this experiment, and subjects of the other two groups participated in "Version B" of this experiment.

Materials

Sixty four word-word pairs were selected that were phonemically and, if written in the same alphabet, graphemically similar. All words were of the type CVCVC. To be considered phonemically similar, they could differ by only one letter. Word pairs were chosen so that this different letter was the middle consonant in 32 pairs and the initial consonant in the remaining 32 pairs. As in Experiment 1, contexts and targets were chosen to be identical with respect to accent position (i.e., syllable), and type (i.e., short rising, short falling, long rising, or long falling), and without any obvious associative or semantic relationship. All words were singular nominative nouns (masculine and feminine), as in Experiment 1. Sixty four pseudoword targets were derived from the word targets by changing the one consonant that was different from the context (so that pseudowords can also be considered to be derived from the contexts).

There were two kinds of pairs that differed in middle consonant and two kinds of pairs that differed in initial consonant (see Table 4). One member of each pair was transcribed in Cyrillic and the other in Roman, so all pairs were graphemically dissimilar. Consequently, for both "initial different" and "middle different" pairs, contexts and targets were: phonemically similar and graphemically dissimilar (Type II, as defined by Table 1); and phonemically dissimilar and graphemically dissimilar (Type IV). The two types of pairs occurred with both word and pseudoword targets to make a total of 8 stimulus groups with 16 stimuli per group.

Table 4. *Examples of context-target pairs used in Experiments 2, 3, and 6, in which members differ either in the initial phoneme or in a phoneme near the middle.*

	Target		Context word	
	Word	Pseudoword	Type II*	Type IV
Initial:				
	KUVAR	GUVAR	ЧУВАР	ЗАЧИН
	NAČIN	FAČIN	ЗАЧИН	ЧУВАР
	ЖИВАЦ	ЂИВАЦ	LIVAC	ČEPIĆ
	РЕПИЋ	ЊЕЛИЋ	ČEPIĆ	LIVAC
Middle:				
	PUŽIĆ	PUDIĆ	ПУТИЋ	РАКУН
	RACUN	RAŽUN	РАКУН	ПУТИЋ
	ЛИКЕР	ЛИГЕР	LIDER	BUKAČ
	БУХАЧ	БУЉАЧ	BUKAČ	LIDER

*Refers to the type of context-target pair identified in Table 1 that is produced by pairing the given context word with the target.

Design

A given context was seen by an individual subject twice, once preceding the word target and once preceding the derived pseudoword target. A given subject never encountered a target word or pseudoword in any of the pairs more than once but every subject saw every type of pair, and every target appeared in every type of pair. These constraints were met with two counterbalancing lists of 128 experimental stimulus pairs. Each list was prepared in two versions (A and B) that were distinguished only by the pseudorandom serial order of the same stimulus pairs: Those stimulus pairs which were adjacent on List A were not adjacent on List B, and vice versa. Thus, the present experiment consisted of "twins" (Version A, and Version B) which were run intermittently. The experimental sequence was preceded by 40 practice pairs.

Procedure

The procedure was the same as in Experiment 1.

Results and Discussion

A four-way analysis of variance (Stimulus Presentation Order by Lexicality by Position by Phonemic Similarity) was computed using subjects as units. Neither the main effect of stimulus presentation order (Version A vs. Version B) nor any interaction with presentation order proved significant. Given this fact, the latency and error data were collapsed over the presentation order variable. Resulting mean latencies, standard deviations, and errors for both words and pseudowords are presented in Table 5. A three-way analysis was conducted on latencies and errors with factors of lexicality, position (initial vs. middle), and phonemic similarity. The position variable is uninterpretable as a main effect. It includes the dissimilar control pairs (Type IV) which, by definition, are not constructed according to the initial different or middle different constraints that govern the phonemically similar pairs (Type II). For the same reason, the interaction of position by lexicality cannot be given a meaningful interpretation. Position by similarity, however, is readily interpreted and is a relation of major interest. Lexicality was significant, *min F'* (1, 112) = 4.47, $p < .05$, similarity was not, *min F'* (1, 107) = 1.73, $p > .05$. Position x similarity was significant, *min F'* (1,107) = 6.27, $p < .01$. For initial different pairs the similarity effect was +36 ms, meaning that the target was facilitated, for middle different pairs the similarity effect was -12 ms, meaning that the target was inhibited. Also interpretable and of major interest is the dependence of the interaction of position by similarity on lexicality, *min F'* (1, 85) = 3.82, $p < .05$. No effects were significant in the error analysis.

The word error and latency data, and the pseudoword error and latency data, were analyzed separately with factors of position and similarity. The only significant outcome was the important interaction in the word latency data between similarity and position (initial different, similarity effect = +55 ms vs. middle different, similarity effect = -27 ms), *min F'* (1,77) = 9.81, $p < .01$.

Table 5. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 2.*

Target	Position of different phoneme	Phonemic relation			
		Similar		Dissimilar	
		L	ER	L	ER
Word:					
	initial	730[a]	3.69	785	7.39
		75[b]	5.34	71	4.15
		70[c]	6.87	88	15.39
	middle	805	5.68	778	4.26
		70	5.07	53	5.59
		87	7.21	76	7.99
Pseudoword:					
	initial	805	4.55	821	5.68
		63	5.17	56	7.19
		67	8.00	66	7.21
	middle	797	2.84	799	2.84
		48	5.70	57	5.36
		63	4.86	76	5.86

[a] Mean

[b] standard deviation by subjects

[c] standard deviation by items

Thus, phonemic similarity effects depend on which parts of the context and target are shared. Sharing all but the initial consonant resulted in a positive similarity effect; sharing all but the middle consonant resulted in a negative similarity effect, as in Experiment 1. The effect of position was more noticeable for words than for pseudowords. The results are consistent with the idea that the activity level of a partially activated word unit depends on which parts of the unit are activated. They are also consistent with the results of Experiment 1 in demonstrating that the observed similarity effects, for either type of phonemic similarity, were produced by shared phonemes and not by orthographic or form similarity. In the present experiment one member of each context-target pair was transcribed in Cyrillic, the other in Roman. In sum, Experiments 1 and 2 suggest that the observed similarity effects with Serbo-Croatian words occurred because phonological information was extracted automatically from the contexts, and that this affected lexical decisions on subsequent phonologically related targets.

The difference between the positive and the negative similarity effects may not, however, be simply a matter of a difference in the position of the unshared phoneme. A negative phonemic similarity effect has been reported by Colombo (1986) for Italian context-target pairs of the type CALMO-PALMO. A negative phonemic similarity effect is said by Colombo (1986) to be consistent with models in which inhibition among word processing units is a major mechanism for selecting the most likely unit among the many physically similar candidates activated by a given word stimulus (e.g., McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). It has been suggested that Colombo obtained a negative phonemic similarity effect rather than a positive phonemic similarity effect because she failed to use a masking stimulus forward of the context (Forster, 1987; Forster, Davis, Schoknecht, & Carter, 1987). A positive phonemic similarity effect is repeatedly demonstrated for orthographically similar English words in the lexical decision task when the context is subject to forward masking (e.g., Forster et al., 1987). It should be emphasized, however, that Colombo (1986) found that the negative phonemic similarity effect was tied to targets of high frequency. Low frequency targets reversed the sign.

EXPERIMENT 3

In this experiment we focus on the dependence of the phonemic similarity effect on the frequency of the target word as reported by Colombo (1986). Lexical decisions on high frequency targets were found to be slower in a rhyming context than in a nonrhyming context; the opposite was the case for low frequency targets. In this third experiment we manipulate target frequency but do so in the form of experiential familiarity (Gernsbacher, 1984). Subjective ratings of word familiarity are obviously more contemporary than printed frequency counts. They are also probably more representative of word experiences given that they take into account words spoken, written and heard. Although such ratings will vary with populations, such variation is immaterial when the ratings are provided by people from the same population as the subjects in the experiment. Such was the case in the present experiment and the subsequent experiments in the present article using familiarity as a variable. The distinct advantage, in experimental terms, of using an experiential familiarity measure is that it is a more reliable predictor of response latency and accuracy in word recognition experiments than printed frequency counts (Gernsbacher, 1984).

Target familiarity was investigated in the present experiment in connection with a further examination of the effect of manipulating the position of the letter (middle or initial) distinguishing a target word from its phonemically similar context.

Method

Subjects

Twenty-two high school seniors from the Fourth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory. After all subjects had been run it was found that one of them was not a native speaker of Serbo-Croatian. This particular subject's data were excluded from analyses.

Materials

Sixty four word-word pairs were selected that were phonemically similar. To be considered phonemically similar, they could differ by only one letter. Word pairs were chosen so that this different letter was the middle consonant in 32 pairs and the initial consonant in the remaining 32 pairs. In addition, word pairs were chosen so that half of them contained targets of high familiarity and half of them contained targets of low familiarity. The words were drawn from the upper and lower ends of a set of 120 words judged for familiarity on a 5 point scale (5 = very familiar, 1 = very unfamiliar) by 35 students from the same year as the high school students who served as subjects. The individual familiarity judgments were summed up for each word at each level of the 5-point scale. As in the previous experiments, contexts and targets were chosen to be identical with respect to accent position and type and without any obvious associative or semantic relationship. As in Experiment 2, all words were singular nominative nouns (masculine and feminine), of two syllables and five letters. Sixty-four pseudoword targets were derived from the word targets by changing the one consonant that was different from the context.

For both "initial different" and "middle different" pairs, as illustrated in Table 4, contexts and targets were: (a) phonemically similar and graphemically dissimilar (Type II); and (b) phonemically and graphemically dissimilar (Type IV). These two types occurred for both word and pseudoword targets.

Design

A given context was seen by an individual subject twice, once preceding the word target and once preceding the derived pseudoword target. A given subject never encountered a target word or pseudoword in any of the pairs more than once but every subject saw every type of pair, and every target appeared in every type of pair. These constraints were met with two counterbalancing groups. In total, each subject saw 128 experimental stimulus pairs ordered randomly. The experimental sequence was preceded by 32 practice pairs of the same types.

Procedure

The procedure was the same as in Experiment 1.

Results and Discussion

Table 6 presents the mean reaction times, standard deviations, and errors for words and pseudowords. An analysis of variance was conducted on all data with factors of lexicality, position, and phonemic similarity. Position as a main effect, and its interaction with lexicality, are not interpretable for the reasons cited in Experiment 2. In the latencies analysis, phonemic similarity was not significant, $min\ F'(1,129) = 1.8$, $p > .05$; lexicality was significant, $min\ F'(1, 104) = 8.0$, $p < .01$. The important interaction between phonemic similarity and position was significant, $min\ F'(1,116) = 6.96$, $p < .01$, (for initial position, similarity effect = +35 ms, for middle position, similarity effect = -9 ms). There was no three way interaction, $min\ F' < 1$. In the error analysis, only lexicality reached significance, $min\ F'(1, 85) = 4.24$, $p < .05$.

To incorporate target familiarity, a second analysis was performed on the word data with factors of target familiarity, phonemic similarity, and position. The relevant data are presented in Table 7.

Table 6. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 3.*

Target	Position of different phoneme	Phonemic relation Similar L	Similar ER	Dissimilar L	Dissimilar ER
Word:					
	initial	689[a]	5.36	731	6.25
		110[b]	5.33	106	7.91
		61[c]	9.64	79	7.30
	middle	730	6.25	718	3.87
		122	7.40	123	4.63
		48	11.67	72	6.29
Pseudoword:					
	initial	743	4.76	771	3.57
		101	7.37	90	4.67
		40	9.57	62	5.77
	middle	745	2.08	739	1.10
		85	3.61	99	2.51
		53	0.00	53	2.50

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

Table 7. *Mean word decision latencies L (in ms) and error rate ER (in %) as a function of target familiarity, with the corresponding standard deviation by subjects and by items (in ms) in Experiment 3.*

Position of different phoneme	Target familiarity	Phonemic relation Similar L	Similar ER	Dissimilar L	Dissimilar ER
Initial:					
	low	706[a]	8.93	787	9.52
		126[b]	9.80	123	11.80
		91[c]	10.28	83	10.72
	high	672	1.79	679	2.98
		105	4.48	101	6.74
		46	5.04	34	4.6
Middle:					
	low	750	8.93	771	4.76
		126	12.59	133	7.37
		50	12.07	40	6.24
	high	712	3.57	664	2.98
		125	5.79	119	6.74
		50	6.92	32	4.62

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

We consider the latencies first. Familiarity was significant (high = 706 ms vs. low = 731 ms), *min F'* (1, 78) = 37.16, $p < .001$ but phonemic similarity was not, *min F'* (1, 77) = 1.61, $p > .05$. The two phonemic similarity interactions proved significant: similarity by familiarity (for high familiarity, similarity effect = -21 ms, for low familiarity, similarity effect = + 51 ms), *min F'* (1, 73) = 9.59, $p < .01$; and phonemic similarity by position (for initial position, similarity effect = +42 ms, for middle position, similarity effect = -12 ms), *min F'* (1, 76) = 6.56, $p < .01$. There was no three-way interaction, *min F'* < 1. The error analysis found only the main effect of familiarity to be significant, *min F'* (1, 55) = 6.93, $p < .01$.

The present results confirm that the effect of phonemic similarity varies as a function of which phonemes are shared. The effect is positive when all phonemes but the initial phoneme are shared and negative when all phonemes but the middle phoneme are shared. The same pattern had been observed in Experiment 2, and the latter aspect had been observed in Experiment 1. In addition, the present results concur with those of Colombo (1986) in showing that the familiarity of the target determines the sign of the phonemic similarity effect; it is negative when the familiarity is high and it is positive when the familiarity is low.

How might the dependence of the phonemic similarity effect on target familiarity be explained? In consensus with Colombo (1986), we believe that the answer lies with activation assumptions that McClelland and Rumelhart (1981, pp. 379 - 381) incorporated into their network model of word recognition. Three assumptions are of relevance. First, the units for high-frequency words have resting levels higher than those for low-frequency words, so that for the same stimulation representations of high-frequency words are activated sooner. Second, inactive units have no influence on their neighbors. Only units in an active state have any effects, either excitatory or inhibitory. Third, the net input to a unit drives the activation of the unit up or down, depending on whether it is positive or negative. The degree of the effect of the input on the unit is modulated by the unit's current activity level to keep the input to the unit from driving it beyond some maximum and minimum values (Grossberg, 1978). This means that the inhibition felt by a word unit is dependent on its level of activation. The latter assumption will be referred to as "Grossberg's principle of self-modulation." Let us apply the preceding assumptions to phonemically similar and dissimilar context-target sequences.

A phonemically similar context word will activate partially a number of word units, namely, those that have some of the same phonemes in the same positions as the context word. One of these partially activated word units represents the target. If the target is a high familiarity word, then the activation level of its representation will be relatively high and, by Grossberg's principle of self-modulation, the amount of inhibition that its representation receives from other units during the processing of the context word will also be relatively high. Consequently, when the target word does occur it is processed in circumstances in which its corresponding word unit is significantly inhibited, thereby slowing the target's recognition. A phonemically dissimilar context, in contrast, would not activate the target's representation. The target's representation would not, therefore, be subject to inhibition prior to the occurrence of the target. In sum, target processing should be slower in phonemically similar contexts.

Colombo's (1986) argument, as we have stated it thus far, predicts that the phonemic similarity effect should always be negative, with the magnitude of this negative effect proportional to a target's familiarity. However, in her work, and in the data of the present experiment, phonemically similar contexts with low familiarity targets produced positive similarity effects. One way to accommodate this result is to require that a given word unit is subject to inhibition only when its activation level exceeds a criterial value (Colombo, 1986). If word units representing low familiarity targets do not reach this criterion, then they will tend to escape inhibition. They will, however, benefit from being activated above the resting level by the occurrence of a phonemically similar context. Consequently low familiarity targets, with all the benefits of activation and none of the costs of inhibition, will be responded to faster in phonemically similar contexts than in phonemically dissimilar contexts. There is a better way, however, to accommodate the facilitation with low familiarity targets, one that dispenses with the need to assume a further specialized mechanism. A phonemically similar context activates levels of units below the word unit level that will be used in the processing of the target. Letter and/or phoneme units to be used in processing the target are primed, and these primed units are sources of activation for the word unit to be used in processing the target. Other things being equal, these activations induced by the phonemically similar context would guarantee a facilitation of the response to the target. Apparently, in the case of high familiarity targets, the inhibition generated at the word unit level exceeds the excitation generated below whereas in the case of low familiarity targets it does not.

In Colombo's (1986) studies, similarity and lexicality tended to interact. In conditions where the sign of the similarity effect was negative for words it was positive for pseudowords. If a negative similarity effect is due strictly to inhibition of word units, then it should not occur for pseudowords. Consider a word context-pseudoword target sequence. A context word produces activation at prelexical levels (e.g., letter, phoneme) as well as at the word unit level. Word units activated to a certain level begin to inhibit other word units. The set of activated word units sends activation to the adjacent prelexical level but no inhibition; in McClelland and Rumelhart's (1981) model, inhibition operates only within a level. That some word units are inhibited at the time of presentation will not affect the speed of processing a target pseudoword. In contrast, the fact that some prelexical units are activated by the context, namely, those that will be involved in the processing of the target pseudoword, can influence the speed of processing a target pseudoword. Accordingly, the phonemic similarity effect for pseudoword targets, like that for low familiarity words, should be positive. In the data of the present experiment, however, the phonemic similarity effect was clearly positive only for initial different pseudowords. For middle different pseudowords the numerical difference between phonemically similar and phonemically dissimilar conditions was small and negative. In Experiment 1 it was also the case that middle different pseudowords failed to behave in the manner expected and observed of low familiarity words. The pseudoword data from Experiment 2 are more compatible with expectations, but they fall short of providing confirmation. We should note, however, that the predicted similarity by lexicality interaction is observed in a subsequent experiment (Experiment 9).

With respect to the position variable, we hypothesized above that the activation level of a partially activated word unit depends on which parts are activated. Specifically, we suggested that a context that differs from a target in the initial letter/phoneme position may activate the target's representation less than a context that differs in the middle letter/phoneme position. The general hypothesis and the specific suggestion can be accommodated by the preceding arguments. Let us return to the example of context-target sequences of three-letter stimuli. Consider the initial different sequence XBC-ABC and the middle different sequence AXC-ABC. The context XBC will activate the word unit ABC to the level of a_i. In comparison, the context AXC will activate the word unit ABC to the level a_m, where $a_m > a_i$ by argument. Because a unit's activation level is a significant determinant of the degree to which the unit is inhibited, then ABC will be inhibited more in the AXC context than in the XBC context. Further, if a_i is sufficiently small so that the inhibition generated from units at the word unit level fails to exceed the excitation generated from units below the word unit level, then the XBC-ABC sequence will produce a positive phonemic similarity effect.

EXPERIMENT 4

In the present experiment we extend to contextual items our study of familiarity's contribution to the phonemic similarity effect. Consider phonemically similar pairs comprising target words of high familiarity. Whether the context is a high familiarity word or a low familiarity word, the target's representation should be inhibited and a negative similarity effect should be observed. Similarly, context familiarity would not be expected to affect the sign of the similarity effect for low familiarity targets. Regardless of context frequency, the word units representing low familiarity words should receive relatively little inhibition. This means that for a high or low familiarity context followed by a low familiarity target, the processing advantages of priming units below the word unit level should outweigh the processing disadvantages of inhibiting the target's word unit. In sum, we expect to see a similarity by target familiarity interaction but not a similarity by context familiarity interaction.

Although context familiarity should not influence the similarity effect, it should be present as a main effect, and it might be expected to interact with target familiarity. The activation level of the representation of a high familiar context will be greater than that of a low familiar context at the time of target presentation. If the lexical decision process is understandable ultimately as a signal detection problem, then decision speed and accuracy will vary inversely with the signal (activity level of target word unit)-to-noise (activity levels of other word units) ratio.

This ratio would be smaller, on the average, for high familiarity contexts. Additionally, the signal-to-noise hypothesis suggests that there should be an interaction of context familiarity and target familiarity: This ratio should be smallest for the combination of high familiarity contexts and low familiarity targets and largest for the combination of low familiarity contexts and high familiarity targets.

The present experiment also extends the investigation of similarity type or, more precisely, the pattern of phonemic similarity between context and target. In the previous experiments a phonemically similar context and target were distinguished by the middle or the initial letter. In the present experiment, context and target in the phonemically similar condition are distinguished by the final letter. We hypothesized above that a context that differs from the target in the final phoneme position may activate the target's representation more than a context that differs in the middle position. Because a unit's activation level, by Grossberg's (1978) self-modulation principle, modulates the degree to which the unit is inhibited, then the target ABC should be inhibited more in the ABX context than in the AXC context. We should expect, therefore, that the ABX-ABC (final different) word sequences will produce a negative phonemic similarity effect.

Method

Subjects

Twenty high school seniors from the Fourth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials

One hundred and twenty five pairs of words were created that differed only in the final consonant letter. Examples are given in Table 8. All words were between 4 and 6 letters in length, a context word always had the same number of phonemes as its corresponding target word. The individual words in these pairs were then presented for familiarity judgements to 35 high school students from the same school and the same year as the students who participated as subjects in the experiment. As before, a 1 to 5 scale was used. Words ranked 1 to 3 on the average were designated as low familiarity, and words ranked 4 to 5 on the average were designated as high familiarity. On the basis of these judgements, 96 pairs were drawn from the original 125 pairs. Twenty four pairs were selected such that the mean familiarity measure for the context items was between 1 and 3, and the mean familiarity measure for the target items was between 4 and 5. This pattern defined the low context-high target group of stimuli. The same procedure was used to arrive at the other three groupings involving word targets: low context-low target, high context-high target, high context-low target. To create the corresponding word-pseudoword pairings, the final letters of the targets were changed. To create the phonemically dissimilar pairs, contexts were interchanged among the original (phonemically similar) pairs within each of the four word-target and four pseudoword-target groupings. There were in all, therefore, 16 groupings of stimuli. All other features of the context-target pairs were as in the preceding experiments.

Table 8. *Examples of context-target pairs used in Experiments 4 and 7 in which members differ in the final phoneme.*

Target		Context word	
Word	Pseudoword	Type II*	Type IV
GUSAN	GUSAD	ГУСАР	ФАКИН
FAKIR	FAKIZ)	ФАКИН	ГУСАР
ПАСУЉ	ПАСУФ	PASUS	DUVAČ
ДУВАН	ДУВАГ	DUVAČ	PASUS

*Refers to the type of context-target pair identified in Table 1 that is produced by pairing the given context word with the target.

Design

The 8 word-word and the 8 word-pseudoword conditions were manipulated within subjects. No subject saw the same target twice, and the targets that appeared in each condition were counterbalanced over subjects. This was achieved for the 8 word-word groupings and the 8 word-pseudoword groupings by reversing context words within pairs of a given grouping of stimuli for half the subjects. Each subject was presented with 12 stimuli pairs in each of the 16 conditions for a total of 192 pairs.

Procedure

The procedure was the same as in Experiment 1.

Results and Discussion

Table 9 presents the mean latencies, standard deviations, and errors for words and pseudowords. Two sets of analyses were conducted. The first was performed on the latencies and errors of both the word and pseudoword data ignoring target familiarity since this variable was undefined for pseudoword targets. The second was limited to the word data.

Table 9. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 4.*

Target	Context familiarity	Phonemic relation			
		Similar		Dissimilar	
		L	ER	L	ER
Word:					
	high	738[a]	11.25	709	8.96
		91[b]	6.77	96	5.28
		100[c]	17.09	106	16.40
	low	697	3.96	683	4.79
		95	6.41	97	5.45
		79	7.07	92	11.11
Pseudoword:					
	high	757	3.12	807	3.13
		92	4.85	102	4.25
		69	7.19	81	6.24
	low	770	3.34	780	3.75
		114	4.99	102	4.66
		86	5.95	90	9.14

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

The factors in the first analysis were lexicality, phonemic similarity, and context familiarity. Lexicality was significant (words = 707 ms vs. pseudowords = 779 ms), *min* F' (1, 62) = 23.27, $p < .001$, as was context familiarity (high = 753 ms vs. low = 733 ms), *min* F' (1,165) = 4.15, $p < .05$. Phonemic similarity was insignificant as a main effect (similar = 741 ms vs. dissimilar = 745 ms), *min* $F' < 1$, but interacted significantly with lexicality (word similarity effect = -22 ms vs. pseudoword similarity effect = +31 ms), *min* F' (1,194) = 7.14, $p < .01$. In contrast, the interaction of similarity with context familiarity was not significant, *min* $F' < 1$. Context familiarity by lexicality also failed to reach significance, *min* F' (1,123) = 2.4, $p > .05$. The error analysis confirmed the insignificance of similarity, *min* $F' < 1$, and the significance of lexicality (words = 7.24% vs pseudowords = 3.33%), *min* F' (1,95) = 7.15, $p < .01$, and of context familiarity (high = 6.61% vs. low = 3.96%), *min* F' (1, 164) = 4.24, $p < .05$. Only one interaction in the error analysis reached significance, namely, lexicality by context familiarity, *min* F' (1, 120) = 4.93, $p < .05$: For words, high = 10.10% and low = 4.38%. For pseudowords, high = 3.13% and low = 3.54%. All other interactions yielded *min* F' values of less than 1.

Because of the interactions with lexicality, a separate analysis was performed on the pseudoword data with the same three factors. These show that the majority of effects observed in the overall analysis are due to the word data (that are addressed in the second set of analyses). Thus, only phonemic similarity proved significant for pseudoword latencies: *min* F' (1,77) = 4.96, $p < .05$.

The word analyses were conducted with factors of similarity, context familiarity, and target familiarity. The data are summarized in Table 10. Considering latencies, the two familiarities were both significant: For context, *min* F' (1,111) = 7.68, $p < .01$; for target, *min* F' (1,84) = 29.41, $p < .001$. Similarity was not significant, *min* F' (1,111) = 2.57, $p > .05$, but its interaction with target familiarity was significant, *min* F' (1,72) = 4.65, $p < .05$. For high familiarity targets, the phonemic similarity effect was negative (similar = 693 ms vs. dissimilar = 644 ms). For low familiarity targets, the similarity effect was numerically positive (similar = 744 ms vs. dissimilar = 752 ms). Notably, neither the interaction of similarity and context familiarity, nor the interaction of context familiarity and target familiarity were significant, *min* $F' < 1$. Considering errors, again both familiarity variables were significant: For context, *min* F' (1,103) = 6.78, $p < .01$: for target, *min* F' (1,107) = 16.24, $p < .001$. Similarity was again insignificant, *min* $F' < 1$. The error analysis did reveal an interaction between context and target familiarity, as can be clearly seen from inspection of Table 10, *min* F' (1,109) = 4.45, $p < .05$. When target familiarity was high, context familiarity did not make a difference to response accuracy (high context = 3.54%, low context = 2.29%), when target familiarity was low, context familiarity affected response accuracy (high context = 16.67%, low context = 6.46%).

Table 10. *Mean word decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 4.*

Word familiarity		Phonemic relation			
Context	Target	Similar		Dissimilar	
		L	ER	L	ER
high	high	712[a]	4.17	658	2.92
		90[b]	6.34	93	4.89
		68[c]	5.84	53	4.64
low	high	674	2.97	631	1.67
		92	6.77	91	4.36
		82	4.64	53	3.81
high	low	768	18.27	764	15.01
		100	10.68	114	9.21
		115	21.40	108	21.26
low	low	721	5.00	741	7.92
		112	7.84	116	8.32
		68	8.85	88	14.74

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

In summary, the results of the present experiment agree largely with expectations. First, whereas the average phonemic similarity effect was negative for words, it was positive for pseudowords. This outcome agrees with the hypothesis that the negative phonemic similarity effect is linked to inhibitory processes acting on word units. Second, the similarity effect with words was negative for high target familiarity (-49 ms) and tended, at least numerically, to be positive for low target familiarity (+8 ms). The manipulation of phonemic similarity in the present experiment was on the final letter (MINUS-MINUT), in contrast to initial and middle letters in preceding experiments. A tentative conclusion is that a difference in final letters, like a difference in middle letters, is more apt to produce a negative phonemic similarity effect than a difference in initial letters. Third, the similarity effect was independent of context familiarity. This was expected on the hypothesis that the inhibition on the target generated in the processing of a phonemically similar context depends on the activation level of the target, which depends significantly on the familiarity of the target.

As expected, high context familiarity slowed responses to the target and produced more errors relative to low context familiarity. This effect of context familiarity was independent of the context-target similarity relation. The outcome was expected based on the assumption that context familiarity exerts its influence on post-lexical processes. The expected interaction of context familiarity and target familiarity was found for errors but not for latencies.

EXPERIMENT 5

For Serbo-Croatian, it has been argued that the pronunciations of printed/written words are based on the states of phoneme processing units that are activated from below by letter processing units and activated from above by word processing units (Lukatela, Turvey, et al., 1989). We wish to develop the argument that, because of this property, phonemic similarity effects with naming should be (a) positive for all types of phoneme similarity, (b) identical for words and pseudowords, and (c) indifferent to the familiarity of the target. The outcome for lexical decision has been that phonemic similarity effects are negative or positive depending on type of similarity and target frequency, and that they may differ for words and pseudowords.

According to our model (Lukatela, Turvey, et al., 1989), there are, in ascending order, feature processing units, letter processing units, phoneme processing units, and word processing units. Within-level interactions are only inhibitory. Between-level interactions are excitatory, and occur in both directions only for adjacent levels. For present purposes, the most significant aspect of the model is the level of phoneme units. One reason is that the assembly of a pronunciation is based on the states of the phoneme processing units. Second, phoneme processing units activate word processing units and are affected, in turn, by the word processing units through feedback connections. In other words, the lexicon is accessed phonologically and not orthographically.

Let us consider what happens to phoneme units and word units in the case of a phonemically similar context and target, for example, DUVAN ("tobacco," Roman)- ДУЋАН ("shop," Cyrillic). The context DUVAN will activate one phoneme unit to a high level at each letter position. Other phoneme units will be activated partially for each letter position. These variously activated phoneme units will activate word units. Many word units will be activated partially. The size of the partially activated set is determined by the number of words that share aspects of DUVAN's phonemic structure. Lateral inhibition trims the set, leaving just those word units that fit most closely DUVAN. Eventually the word unit representing DUVAN will dominate, with the other candidate word units, including that representing the high frequency дУЋАН, inhibited. As this competition among the word units takes place, the phoneme units receive downward activation from the word units. Gradually, the phoneme units most activated from below (letter unit level) will be the ones most activated from above (word unit level). If naming is based on the states of phoneme units, then naming дУЋАН will benefit from the prior occurrence of DUVAN; DUVAN primes all but one of the phoneme units from which дУЋАН 's name is constructed. In contrast, if lexical decision is based on the states of word units, then lexically evaluating the high frequency word дУЋАН should suffer from the prior occurrence of DUVAN; DUVAN inhibits the word unit whose activity level eventually governs the decision that дУЋАН is a word.

Models of English word recognition often exclude a level of phoneme processing units (e.g., Aaronson & Ferres, 1983; Baron, 1973; Becker, 1976, 1979; Kolers, 1970; Paap, Newsome, McDonald, & Schvaneveldt, 1982; Smith, 1971). The idea is that the phonology of English words is derived from the lexicon and not pre-lexically, as advocated above for Serbo-Croatian words. A model that excludes phoneme level units makes different predictions about the phonemic similarity effect for naming. Most notably, it predicts that the similarity effect for naming should be qualitatively similar to the phonemic similarity effect for lexical decision. Negative similarity effects should be observed, target familiarity should affect the sign of the effect, and the similarity effect may differ for words and pseudowords. These predictions follow largely from the fact that, for a model without phoneme processing units, the activity levels of word units govern the speed of naming. Thus, the familiarity of a target in a phonemically similar context will determine whether or not the activation level of its lexical representation at the time of target presentation is above or below the resting level.

The present experiment examines the phonemic similarity effect in naming for context-target sequences that differ in the middle letter. It is basically a replication of Experiment 1 with the exception of the dependent measure. Our model leads us to expect a positive phonemic similarity effect for both words and pseudowords, and independence of the effect from graphemic similarity.

Method

Subjects

Fifty-two high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials and Design

These were the same as Experiment 1.

Procedure

Subjects were required to pronounce each target letter string as quickly and distinctly as possible. In all conditions, latencies from the onset of the target to the onset of the response were measured by a voice operated trigger relay. Naming was considered to be erroneous when the pronunciation included a phoneme not specified by the characters in the letter string, the pronunciation was not smooth (i.e., the subject hesitated after beginning the name), or the response was not initiated within the cut-off latency. All presentation parameters and cut-off latencies were the same as in Experiment 1.

Results and Discussion

The mean latencies, standard deviations, and errors for words and pseudowords are presented in Table 11. By inspection of Table 11, phonemic similarity was facilitatory of both word and pseudoword naming times independent of graphemic similarity. An analysis of variance was conducted on the latency data with factors of lexicality, phonemic similarity, and alphabet (same vs. other). Both similarity (similar = 668 ms vs. dissimilar = 690 ms) and lexicality (words = 652 ms vs. nonwords = 706 ms) were significant: *min* $F'(1,176) = 11.67$, $p < .001$ and *min* $F'(1,187) = 38.64$, $p < .001$. Alphabet was not significant, *min* $F' < 1$, and the important interactions of similarity by alphabet *min* $F' = 1$, and similarity by lexicality were not significant, *min* $F' < 1$. Also insignificant was the alphabet by lexicality interaction, *min* $F' < 1$. Repeating the analysis for errors produced no effects with all *min* F' values less than 1.

Table 11. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 5.*

		Phonemic relation			
Target	Alphabet	Similar		Dissimilar	
		L	ER	L	ER
Word:					
	same	647[a]	1.82	662	2.03
		89[b]	3.77	87	3.97
		63[c]	3.53	58	3.65
	other	641	1.39	659	3.10
		82	2.88	91	4.18
		56	3.73	61	4.77
Pseudoword:					
	same	687	1.60	711	2.56
		109	2.98	126	3.56
		59	4.27	74	5.16
	other	698	2.78	727	2.78
		102	4.19	114	3.33
		70	5.82	69	4.13

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

EXPERIMENT 6

The novelty of Experiment 5, and the contrary nature of its results with respect to those of the preceding lexical decision experiments, encouraged a replication. The sixth experiment repeats the design of Experiment 5 with the additional manipulation of the type of phonemic similarity. It will be recalled that in the lexical decision experiments similarity effects depended on the position of the different phoneme. These position-dependent outcomes were attributed to processes at the level of word units. According to our model of Serbo-Croatian word recognition, they should not be present in rapid naming.

Method

Subjects

Forty-four high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials

Sixty-four word-word pairs were chosen with the same restrictions as Experiment 1 except that half differed by the first consonant and half differed by (approximately) the middle consonant (see Table 1). Sixty-four word-pseudoword pairs were generated from word targets in the standard way. The experimental sequence was divided into quarters, each preceded by four practice trials of a similar sort following a brief rest. Context-target pairs were ordered randomly within each quarter.

Design

A given subject saw 128 experimental and 16 practice trials. Four lists were counterbalanced as in Experiment 1.

Procedure

Same as Experiment 5.

Results and Discussion

A four-way analysis of variance (Position by Lexicality by Phonemic Similarity by Alphabet) was computed using subjects as units. Neither the main effect of position (initial vs. middle) nor any interaction with position proved significant. Given this fact, the latency and error data were collapsed over the position variable. Resulting mean latencies, standard deviations and errors for words and pseudowords are presented in Table 12. By inspection of Table 12, phonemic similarity was facilitatory of both word and pseudoword naming times independent of graphemic similarity. An analysis of variance was conducted on the latency data with factors of lexicality, phonemic similarity, and alphabet. Both similarity (similar = 589 ms vs. dissimilar = 605 ms) and lexicality (words = 576 ms vs. nonwords = 619 ms) were significant: $min\ F'(1,149) = 8.80$, $p < .01$ and $min\ F'(1,169) = 32.57$, $p < .001$. Importantly, neither alphabet nor the interaction of similarity by lexicality were significant, $min\ F'(1, 164) = 3.18$, $p > .05$, and $min\ F' < 1$, respectively. Repeating the analysis for errors produced one significant effect, namely, lexicality, $min\ F'(1,116) = 15.94$, $p < .001$. For all other effects, $min\ F' < 1$. In sum, the present experiment replicated the results of Experiment 6: With rapid naming as the response, the phonemic similarity effect is positive for both words and pseudowords, and was indifferent to graphemic similarity. In addition, the effect of type of phonemic similarity found in the previous lexical decision experiments was not found in the present naming experiment. Results suggest that the naming of Serbo-Croatian words is based on the states of processing units that are prior to word units (prelexical) and subsequent to letter units. In accordance with our model (Lukatela, Turvey, et al., 1989), naming depends on the states of phoneme processing units. The next experiment provides further verification of this claim.

Table 12. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 6.*

Target	Alphabet of context	Phonemic relation Similar L	Similar ER	Dissimilar L	Dissimilar ER
Word:					
	same	567[a]	3.69	580	2.41
		74[b]	4.73	79	3.62
		49[c]	5.30	56	5.43
	other	571	3.35	587	3.69
		77	4.35	79	5.44
		53	6.19	43	6.42
Pseudoword:					
	same	605	6.68	620	7.39
		86	7.67	93	6.91
		48	7.81	54	9.70
	other	614	6.53	636	7.67
		90	7.38	106	6.85
		55	9.37	68	10.27

[a] Mean

[b] standard deviation by subjects

[c] standard deviation by items

EXPERIMENT 7

The seventh experiment examines the contribution of context familiarity and target familiarity to the phonemic similarity effect in naming. In addition, it examines the significance of phonemic similarity defined by distinguishing context and target stimuli by the final letter. The seventh experiment is a naming replication of Experiment 4 in which lexical decision was the dependent measure. In Experiment 4, phonemic similarity interacted with lexicality and with target familiarity but not with context familiarity. The two familiarities were significant main effects in the fourth experiment but an interaction between them was only evident in the error analysis. Our model predicts that, with naming speed and accuracy as the dependent measures, the phonemic similarity effect should be (a) positive for both words and pseudowords, and (b) independent of both target familiarity and context familiarity. We expected that (c) there should be no main effect of context familiarity.

In lexical decision, the arguments marshaled by Colombo (1986) and developed above in reference to Serbo-Croatian language materials, identify target frequency or familiarity as the major determinant of the sign of the phonemic similarity effect. We proposed that a phonemically similar context will partially activate the word unit of the target because of phonemic similarity, and that the target's word unit will become inhibited if the target is a high frequency (familiarity) word and escape inhibition if the target is a low frequency (familiarity) word. As a consequence, the effect of phonemic similarity in lexical decision is negative for high frequency words and positive for low frequency words.

The account of naming in Serbo-Croatian claims that naming is based on the states of the phoneme processing units and not on the states of the word processing units. To repeat the example of DUVAN-дУЋАН from above, although the word unit of the high familiarity word дУЋАН (meaning "shop") is inhibited in the processing of DUVAN, the phoneme units corresponding to the letters д, У, А, and Н are highly activated. The same degree of phoneme unit facilitation would be expected if the target in a phonemically similar context-target sequence was a low familiarity word (or a pseudoword). Consequently, if naming is based on the states of the phoneme processing units, then target familiarity should not affect the similarity effect observed when words are named.

Let us now turn to context familiarity. The main effect of context familiarity in lexical decision observed in Experiment 4 was attributed to post-lexical processes. Relative to a context word of low familiarity, a context word that is of high familiarity affects the post-lexical decision process by reducing the signal-to-noise ratio given by the target's activity level relative to the level of other highly active units. This post-lexical binary decision process based on relative signal strength is absent from naming (Carello et al., 1988; Seidenberg et al., 1984). Consequently, no main effect of context familiarity is expected. An interaction between target familiarity and context familiarity would follow from the fact that the signal-to-noise ratio varies with the familiarity of the target relative to the familiarity of the context. This interaction expected for lexical decision is not expected for naming, again for the reason that naming does not involve post-lexical decision processes.

Method

Subjects

Twenty-two high school seniors from the Fourth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials and Design

Same as in Experiment 4.

Procedure

Same as in Experiment 5.

Results and Discussion

Table 13 presents the mean latencies, standard deviations and errors for words and pseudowords. Two sets of analyses were conducted. The first was performed on the latencies and errors of both the word and pseudoword data. Target familiarity was ignored since this variable was undefined for pseudoword targets. The second was limited to the word data.

Table 13. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 7.*

Target	Context familiarity	Similar L	Similar ER	Dissimilar L	Dissimilar ER
Word:					
	high	579[a]	2.65	605	3.60
		65[b]	2.42	74	3.47
		45[c]	4.95	55	6.95
	low	588	1.89	620	2.08
		75	2.48	79	3.58
		42	4.18	44	4.29
Pseudoword:					
	high	628	3.79	648	7.58
		82	3.84	99	6.38
		60	7.46	65	12.06
	low	624	3.41	668	6.44
		82	3.99	96	4.01
		50	5.51	53	7.25

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

The analyses performed on all data had factors of lexicality, phonemic similarity, and context familiarity. For both latencies and errors, lexicality proved significant, $min\ F'$ (1,102) = 32.78, $p < .001$, and $min\ F'$ (1,77) = 7.64, $p < .01$, respectively. Similarly, for both latencies and errors, phonemic similarity proved significant, $min\ F'$ (1,68) = 18.05, $p < .001$ (similar = 605 ms vs. dissimilar = 635 ms), and $min\ F'$ (1,112) = 4.70, $p < .05$ (similar = 2.94% vs. dissimilar = 4.92%), respectively. The main effect of context familiarity was not significant for either latencies or errors. Considering latencies, only the phonemic similarity by context familiarity interaction exceeded a $min\ F'$ value of 1, $min\ F'$ (1,138) = 1.62, $p > .05$. In all other cases, for both latencies and errors, interactions proved nonexisting, $min\ F' < 1$.

The word analysis had factors of phonemic similarity, context familiarity, and target familiarity. Table 14 summarizes the data. Considering latencies, the main effects of phonemic similarity and target familiarity were significant. For phonemic similarity (similar = 583 ms vs. dissimilar = 613 ms), $min\ F'$ (1, 86) = 12.95, $p < .001$; for target familiarity (high = 590 ms vs. low = 606 ms), $min\ F'$ (1,112) = 4.73, $p < .05$. Context familiarity was insignificant, $min\ F'$ (1,110) = 2.31, $p > .05$. There were no interactions. In all cases, $min\ F' < 1$. Considering errors, there were no main effects and no interactions (only the context familiarity by target familiarity exceeded a $min\ F'$ value of 1, $min\ F'$ (1,94) = 1.92, $p > .05$).

Table 14. *Mean word naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 7.*

Word familiarity Context	Word familiarity Target	Similar L	Similar ER	Dissimilar L	Dissimilar ER
high	high	574[a]	2.27	598	1.52
		64[b]	3.80	78	3.29
		44[c]	4.83	53	3.46
low	high	582	1.89	606	2.27
		78	3.57	85	4.59
		44	3.77	43	4.02
high	low	584	3.03	612	5.68
		68	4.10	75	6.99
		46	5.13	58	8.81
low	low	594	1.89	635	1.89
		74	3.57	82	3.57
		40	4.63	41	4.63

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

We predicted that the effect of phonemic similarity would be the same for both words and pseudowords, as observed in Experiments 5 and 6. This prediction was satisfied in the present experiment. More importantly, we predicted that the similarity effect would be indifferent to target familiarity and context familiarity, and that context familiarity would not be a significant main effect. These predictions were also upheld. The results imply that naming Serbo-Croatian letter strings (a) is largely immune to the actual states of word processing units and the inhibitory relations among them, and (b) does not engage post-lexical processes. Point (a) does not deny lexical factors in naming. Rather, it underscores the view that the word unit level is not the level of fundamental significance to naming.

EXPERIMENT 8

The data of Experiments 5, 6 and 7 support the central role of phoneme processing units in assembling the pronunciation of a Serbo-Croatian word or pseudoword (Lukatela, Turvey, et al., 1989). The states of the phoneme processing units depend upon activation from the letter processing units below and activation from the word processing units above. We now identify two features of the Serbo-Croatian language that reflect an important contribution of word processing units to the naming of words. As noted in the introduction, there are words whose pronunciations cannot be derived unequivocally from the letter-to-phoneme correspondencies. Given the presence of letters that are assigned different phonemes in the two alphabets, the automatically realized connections between individual letters and individual phonemes will yield a number of possible pronunciations at the phoneme unit level. The level of word units contains the information needed to specify uniquely the pronunciation of phonologically ambiguous words. Within the proposed model for Serbo-Croatian word perception and naming, the equivocality is resolved through the automatic two-way interaction of phoneme processing units and word processing units (Lukatela, Turvey, et al., 1989).

The second feature of the Serbo-Croatian language implicating a significant role for word processing units in naming is the fact that the stress pattern is not marked in the orthography. The majority of Serbo-Croatian words display a standard accentuation pattern with stress on the first syllable. The minority are stressed on the second syllable. Moreover, the stress can be either long or short, rising or falling. Pronouncing a Serbo-Croatian word requires knowledge of its stress pattern. That knowledge is not given through the letter units-to-phoneme units connections. It is only made available by accessing word processing units. Below, we suggest how the model for Serbo-Croatian word perception can be modified to incorporate stress information in naming.

In Experiments 1 through 7, the context and target words were phonologically unambiguous. All were stressed on the first syllable and the stress was of the same type in both context and target. Consider one such sequence, the context BALON (balloon) and the target BARON (baron). The phoneme units /b/, /a/, /o/, and /n/ shared by both words would be activated by BALON's presentation. These aspects were spelled out above. We now add the assumption that phoneme units that receive and are activated by stimulation are connected to phoneme units that guide articulation, and that the connections between the two kinds of phoneme units are realized automatically: If the input phoneme unit /b/ is activated (above threshold) then the output phoneme unit /b/ will be activated similarly. As a consequence of the aforementioned connections, BALON's presentation primes the articulation of BARON by priming most of the required output phoneme units that are connected, in turn, to units of articulation (cf. Jordan, 1986).

There is a further way in which we assume that BALON primes BARON's pronunciation. BALON activates its word unit and, thereby, activates information about its stress pattern. This information propagates from the word unit down to the articulatory units activated by BALON's phonemic constituents. As a result, a stress pattern appropriate to the pronunciation of BARON is defined over articulatory units that are similar in kind to those comprising the pronunciation of BARON (specifically, the articulatory unit for /a/ is assigned a short accent). In addition to this short term articulatory encoding bias, there is also a long term articulatory encoding bias. Because of the strong bias in the language to first syllable stress, it can be assumed that the assembly of a pronunciation is biased similarly. Stress on the first syllable is the default option.

In the present experiment, the stress pattern of the context and target never concur. They disagree on place of stress and/or manner of stress (long vs. short, rising vs. falling). In half of the sequences, as in Experiments 5, 6 and 7, the context and target are phonemically similar letter strings. Thus, with respect to assembling the target's pronunciation, the context will activate (a) the appropriate output phonemes and inappropriate stress on one half of the trials, and (b) irrelevant output phonemes and inappropriate stress on the other half of the trials. According to the arguments presented above, in situation (a) pronouncing the target requires changing the stress pattern on a previously activated articulatory unit. In situation (b), pronouncing the target requires activating the appropriate articulatory units and assigning the appropriate stress. If changing the stress pattern consumes much more time than is gained by sensitizing articulatory units, then naming in situation (a) should be slower than naming in situation (b); if changing stress consumes much less time than is gained through priming, then (a) should be faster.

Method

Subjects

Fifty-two high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials

Fifty-six word-word pairs of types I, II, III, and IV (types are defined by Table 1) were chosen as in Experiment 1 but this time they were chosen to be different in either type of accent (rising or falling, long or short) or location of accent (first or second syllable). Examples are given in Table 15. Words could be of two, three, or four syllables. They comprised 6.25 phonemes on average (compared to 5.38 phonemes on average in Experiment 1). Fifty-six word-pseudoword pairs were constructed as detailed previously.

Design

Each subject saw 112 experimental and 32 practice pairs (8 before each quarter of the experimental list), counterbalanced as in Experiment 1.

Procedure

Same as Experiment 4.

Results and Discussion

The data are shown in Table 16. An analysis of the latency data using the factors of phonemic similarity, alphabet, and lexicality found only a main effect of lexicality, *min F′* (1,147) = 30.65, *p* <.001. The main effects of phonemic similarity and alphabet were insignificant, as were all interactions, *min F′* < 1. The same was true of the error analysis: For lexicality, *min F′* (1,156) = 12.0, *p* <.001, and for all other main effects and interactions, *min F′* < 1.

A comparison of Tables 11 and 12 with Table 16 shows that nonuniformity in the stress patterns of context and target words eliminated the phonemic similarity effect in naming. In terms of the hypotheses advanced above, changing stress (either its length or its position) consumed more time than was gained by priming the appropriate articulatory units. For example, if MALINA (raspberry) was the context, with short, rising stress on the first /a/ and MAŠINA (machine) was the target, with long, rising stress on /i/, then (a) the articulatory units of relevance to MAŠINA, namely, those corresponding to /m/, /a/, /i/, /n/, and /a/ in 1st, 2nd, 4th, 5th, and 6th positions, respectively, would be primed, and (b) /a/ in the 2nd position would be stressed. The data suggest that the benefits of (a) are lost by the requirements to shift stress from /a/ in the 2nd position to /i/ in the 4th position, and to change the stress type from short to long.

Table 15. *Examples of context-target pairs used in Experiments 8 and 9 in which members differ in stress pattern.*

Target		Context word			
Word	Pseudoword	Type Ib*	Type III	Type IIb	Type IV
MAŠINA	MAZINA	MALINA	GATARA	МАЛИНА	ГАТАРА
GITARA	GETARA	GATARA	MALINA	ГАТАРА	МАЛИНА
БУЉИНА	ЊУЉИНА	ДУЉИНА	КОЛИЦА	DULJINA	KOLICA
ПОЛИЦА	ЖОЛИЦА	КОЛИЦА	ДУЉИНА	KOLICA	DULJINA

*Refers to the type of context-target pair identified in Table 1 that is produced by pairing the given context word with the target.
Note: Stressed vowels are underlined.

Table 16. *Mean naming latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 8.*

Target	Alphabet	Phonemic relation			
		Similar		Dissimilar	
		L	ER	L	ER
Word:					
	same	624[a]	16.81	603	11.12
		77[b]	11.39	63	8.88
		71[c]	17.67	61	14.59
	other	621	16.54	620	13.80
		64	10.15	76	11.95
		59	17.69	59	15.96
Pseudoword:					
	same	669	6.46	675	9.75
		72	7.11	78	9.39
		66	8.14	84	13.32
	other	681	10.22	676	9.89
		84	8.17	77	11.85
		88	11.84	85	12.68

[a] Mean
[b] standard deviation by subjects
[c] standard deviation by items

The mixing of first- and second-syllable stressed contexts affected the relation between phonemic similarity and the latency to name pseudowords. How is the result with pseudowords to be understood? When the pseuodoword MAZINA is presented, it will activate partially a number of word units, including MALINA and MAŠINA. Because no word unit is fully activated, and because those that are activated partially are stressed either on the first or second syllable, competition will ensue among word units with respect to the assignment of stress to MAZINA. The nature and time course of this competition would be the same for both phonemically similar and dissimilar contexts. Assuming that the time to assemble a pronunciation is scaled, in part, to the time to determine a stress pattern, then it can be supposed that the delay induced by the competitive process is enough to eliminate the advantage of phonemic similarity. That is, even though the articulatory units for MAZINA are prepared faster in a phonemically similar context than in a phonemically dissimilar context, the resolution of stress assignment takes the same amount of time in both contexts. If activating articulatory units and assigning them intensity weights are independent, parallel activities, then a slow assignment will override a swift activation. Importantly, the foregoing situation should not arise in the conditions of Experiments 5, 6 and 7. There, the pseudowords would activate word units that would be uniform in their stress pattern (accented first syllable). Consequently, no competition with regard to stress assignment would ensue.

In approximately 80% of the phonemically similar pairs, the names given to pseudoword targets were stressed as their contexts. In the phonemically dissimilar pairs, target stress was not systematically related to context stress. One implication of these facts is that the competition at the word level was generally unresolved, with the stress assignment going by default, in the phonemically similar situation, to the stress determined by the immediately preceding context. In the phonemically dissimilar situations, the stress of the context plays no major, systematic role in the assembling of the target's pronunciation. It appears that stress is defined with respect to particular articulatory units. If the context and target do not share these units, then the stress pattern of the context is irrelevant to the eventual pronunciation of the target.

One other feature of the pseudoword data is valuable to understanding the naming of Serbo-Croatian words. We divided the target stimuli in the phonemically similar pairs into those preceded by a context stressed on the first syllable (the standard way) and those preceded by a context stressed on the second syllable. In first-syllable contexts the target latency was 669 ms on the average; in second-syllable contexts the latency was 700 ms on the average. This difference is consistent with our hypothesis that the assembling of a pronunciation "prefers" the standard stress pattern.

EXPERIMENT 9

In our model of pronunciation assembly for Serbo-Croatian words, activated word units inform articulatory units about accent via a connective network. The incongruity of context and target in stress pattern is an effect at the level of articulatory units. In our view, lexical decision is based on the states of the word units and involves a post-lexical decision that is constrained by the target unit's activity level relative to the level of other activated units. Thus, lexical decision does not engage the level of articulatory units. A lexical decision replication of Experiment 8 was conducted, therefore, with the expectation that stress incongruity between context and target should not affect the relation between phonemic similarity and decision latency. The experiment uses context-target sequences that differ in middle phonemes. The results of this ninth experiment, therefore, should be the same as the results of the first three experiments: A negative phonemic similarity effect should prevail for words and a less negative, or even a positive, phonemic similarity effect should prevail for pseudowords regardless of the context-target stress relationship.

Method

Subjects

Forty-eight high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials, Design, and Procedure

Same as Experiment 8 with the exception that targets were lexically evaluated instead of named.

Results and Discussion

Table 17 presents the mean latencies, standard deviations, and errors for words and pseudowords. Analysis of variance on the latency data revealed that the main effect of lexicality was significant, $min\ F'$ (1,116) = 29.28, $p < .001$, and that only the interaction between phonemic similarity and lexicality was significant, $min\ F'$ (1,149) = 6.09, $p < .01$. For words the phonemic similarity effect was -24 ms, for pseudowords the similarity effect was +12 ms. The error analyses showed the significance of lexicality, $min\ F'$ (1,139) = 5.43, $p < .05$; the errors were committed more to words than pseudowords. The interaction between phonemic similarity and lexicality proved significant, $min\ F'$ (1,147) = 4.06, $p < .05$. For words the phonemic similarity effect in terms of errors was -2.16%, for pseudowords the phonemic similarity effect in terms of errors was 2.82%. In sum, responses to words were slower and less accurate in phonemically similar pairs than in phonemically dissimilar pairs, with the opposite holding true for pseudowords. This outcome replicates the similarity by lexicality interaction reported by Colombo (1986) and necessitated by the claim that the negative similarity effect is due to interactions at the word unit level. Further, the present experiment shows that the differently signed phonemic similarity effects for words and pseudowords observed for lexical decision when context and target are stressed identically, are preserved when context and target are stressed differently. This outcome contrasts with that found for naming. The contrasting sensitivity to stress congruity can be attributed to the dominant roles played by phoneme and articulatory units in naming and by word units in lexical decision.

Table 17. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 9.*

Target	Alphabet	Phonemic relation Similar L	 ER	 Dissimilar L	 ER
Word:					
	same	782[a]	9.88	754	8.33
		100[b]	9.05	83	7.98
		75[c]	9.70	72	11.01
	other	790	10.63	771	6.99
		92	9.32	87	6.34
		81	12.36	72	10.98
Pseudoword:					
	same	823	4.76	839	8.93
		104	6.81	101	9.94
		60	6.13	68	11.11
	other	827	3.87	835	5.36
		103	5.70	98	7.29
		71	5.72	74	7.85

[a] Mean

[b] standard deviation by subjects

[c] standard deviation by items

EXPERIMENT 10

The final experiment seeks stronger confirmation of the claim that the phonemic similarity effect in lexical decision is due to the automatic, prelexical activation of phonological information. The experiment has three major features. First, phonemically similar contexts are presented very briefly under forward masking conditions. This is done to minimize the explicit identification of the context but to still allow the context to influence the processing of the target. Such conditions may help to prevent intentional, strategic use of context information (e.g., Evett & Humphreys, 1981; Humphreys, Evett, & Taylor, 1982). Second, each context differs from its target in alphabet and case. For example, if the target is in Cyrillic upper case, then the context is in Roman lower case. This manipulation exaggerates the graphemic dissimilarity between phonemically similar words. Third, pseudoword phonemically similar contexts are used. The reason for their incorporation is explained below.

Consider the argument from the introduction that there are two ways in which a word's phonology can be derived (Coltheart, 1978). One way is through the lexicon. Once a word unit has been activated on the basis of its orthographic properties, phonological information about the word is made available. The other way is through GPCs. Suppose that the lexical way is mandatory (automatic) and that the GPCs way is optional (nonautomatic), as has been frequently proposed (e.g., Coltheart et al., 1977; Henderson, 1982). Then, under conditions that minimize optional strategies, only the lexical way should work. On the basis of the preceding argument Humphreys et al. (1982, p. 581) hypothesized that: "If phonological information is activated via a nonlexical route, a pseudohomophone priming effect should occur." Finding that the pseudohomophone condition (tial-TILE) in their masking experiment did not differ from its graphemic control condition (tirl-TILE), Humphreys et al. (1982) concluded that the lexical way of deriving phonology (a) is the only one of the two that is automatic and (b) may be, in fact, the only route.

The present experiment used a subset of stimuli from Experiment 2 and 3 in which context and target differ in the initial letter (see Table 4). In Experiment 2 and 3, the phonemic similarity effect for these initial different stimuli was positive for both word targets (Experiment 2, +49 ms, Experiment 3, +42 ms) and pseudoword targets (Experiment 2, +30 ms, Experiment 3, +28 ms). The experiment, therefore, used phonemic similarity conditions known to produce facilitation. Our objective was to create experimental conditions which would be comparable, at least in part, to those in the Humphreys et al. (1982) studies of the phonemic similarity effect in English. On the basis of our account of Serbo-Croatian word processing we expect that, in the present experiment, a pseudoword context that is phonemically similar to its target should facilitate lexical decision on the target. Further, we expect that the phonemic similarity effect in pseudoword-word conditions should not differ from that in word-pseudoword conditions. If the phonemic similarity effect in the lexical decision task with graphemically dissimilar stimuli is due primarily to an automatic, prelexical computation of a letter string's phonology, then the phonemic similarity effect should not be affected by the lexical composition of the phonemically similar pair.

Method

Subjects

Forty four high school seniors from the Fifth Belgrade Gymnasium were paid to serve in the experiment. Each subject was assigned to one of four counterbalancing groups according to his or her appearance at the laboratory.

Materials

Examples of the primary stimulus set are to be found in Table 4. The basic description of the types of words used in Experiment 2 applies to the words of the present experiment. As depicted in Table 4, a context and its target differed in the initial phoneme. However, all contexts were presented in lower case, and all targets were presented in upper case. The context and target in a pair never appeared in the same alphabet. In half of the pairs the context was Cyrillic and in half it was Roman. There were 32 such word-word pairs. By changing the initial letter in the contexts, 32 pseudoword-word pairs were created; by changing the initial letter in the targets, 32 word-pseudoword pairs were created; by changing the initial letters in both contexts and targets, 32 pseudoword-pseudoword pairs were created. Phonemic similarity was manipulated only in the 32 pseudoword-word pairs (16 similar, 16 dissimilar) and the 32 word-pseudoword pairs (16 similar, 16 dissimilar). The remaining 64 pairs provided filler items.

Design

There were four groups of subjects, 11 subjects per group. Each subject saw 16 pseudoword (PW)-word (W) rhyming pairs, 16 W-PW rhyming pairs, 16 PW-W control pairs, and 16 W-PW control pairs. A given context and a given target was seen by an individual subject only once. This was achieved as follows. Let A-B, C-D, E-F, and G-H be four rhyming word pairs. For group 1, these pairs are rearranged and individual members modified as follows: A*-B (phonemic similarity), A-D* (phonemic dissimilarity), E-F* (phonemic similarity), E*-H (phonemic dissimilarity), where the asterisk identifies the pseudoword derived

from the designated word by changing the initial letter. For group 2: C*-B (phonemic dissimilarity), C*-D (phonemic similarity), G-F* (phonemic dissimilarity), G-H* (phonemic similarity), and so on for groups 3 and 4. The foregoing procedure was repeated for all 16 subsets of four word-word pairs that made up the set of 64 word-word pairs. As a consequence each subject in each group saw 16 pairs of each of the following: W-PW (phonemic similarity), W-PW (phonemic dissimilarity), PW-W (phonemic similarity) and PW-W (phonemic dissimilarity). The experimental sequences were preceded by 24 practice pairs of the same type.

Procedure

The procedure was the same as previous experiments with the following exception. The fixation point was followed by a forward mask, centered on the screen, consisting of a row of ten hash marks for 500 ms. This was followed immediately by the context for 80 ms and, immediately after that, the target for 500 ms, both in the same location as the mask had been. A pretest with a different group of subjects had shown that the prime could not be reliably identified under these circumstances.

Results and Discussion

Table 18 presents the mean latencies, standard deviations, and errors for the word and pseudoword targets in the conditions in which phonemic similarity was manipulated. As can be seen from inspection of the table, the phonemic similarity effect occurred for both sequence types. A 2 (PW-W sequence vs. W-PW sequence) by 2 (phonemic similarity vs. phonemic dissimilarity) analysis of variance was conducted. The main effects of sequence (PW-W = 765 ms vs. W-PW = 814 ms) and phonemic similarity (phonemic similarity = 774 ms vs. phonemic dissimilarity = 805 ms) were significant, *min F'* $(1,103) = 13.11$, $p < .001$, and *min F'* $(1,102) = 9.08$, $p < .01$, respectively. The important interaction of phonemic similarity by sequence was not significant, *min F'* < 1. The sequence difference should be attributed primarily to the fact that words are accepted faster than pseudowords are rejected. Analysis of errors found only the sequence variable to be significant, with more errors produced on word targets (5.97%) than pseudoword targets (2.06%), *min F'* $(1,105) = 9.58$, $p < .01$. The conclusion to be drawn is that the phonemic similarity effect in lexical decision is due to the automatic, prelexical activation of phonological information.

Table 18. *Mean decision latencies L (in ms) and error rate ER (in %), with the corresponding standard deviation by subjects and by items (in ms) in Experiment 10 with masked, graphemically dissimilar contexts.*

Context	Target	Phonemic relation: Similar L	Similar ER	Dissimilar L	Dissimilar ER
Pseudoword	Word	748[a]	6.25	776	5.68
		82[b]	6.74	81	6.15
		87[c]	8.24	81	9.79
Word	Pseudoword	794	1.70	832	2.41
		96	3.66	93	3.86
		77	5.58	82	6.73

GENERAL DISCUSSION

The present series of experiments used a paradigm in which two Serbo-Croatian words presented sequentially were phonemically similar or dissimilar and graphemically similar or dissimilar. Graphemic similarity was manipulated through the use of the two Serbo-Croatian alphabets. The second word was evaluated lexically or named. Comparisons of latencies and errors were made across stimulus pairs of which the following are typical: PUTIĆ-PU̯ŽIĆ, similar phonemically and similar graphemically; ПУТИЋ-PU̯ŽIĆ, similar phonemically, dissimilar graphemically; RAKUN-PU̯ŽIĆ, a control in which context is in the same alphabet as the target; РАКУН-PU̯ŽIĆ, a control in which context is in a different alphabet from the target. Experiments 1, 2, 3, 4 and 9 showed that the visual phonemic similarity effect in lexical decision does not depend on graphemic similarity. Experiments 5, 6 and 7 showed that the visual phonemic similarity effect in naming does not depend on graphemic similarity. For both lexical decision and naming, the phonemic similarity effect arises from the use in target processing of phonological information activated in context processing. Experiment 10 showed that this activation occurs under conditions that minimize the identification of the context. The activation of the phonological information underwriting the similarity effect appears to be automatic.

We have presented a view of the processing of Serbo-Croatian letter strings as involving, in bottom-up succession, connected layers of featural processing units, letter processing units, phoneme processing units, and word processing units

(Lukatela, Turvey et al., 1989). To incorporate naming, a further layer of articulatory units is required. Of special significance are the letter units-to-phoneme units connections. These embody the grapheme-phoneme correspondences of the language. In the Serbo-Croatian orthography, unlike the English orthography, these correspondences are straightforward. All letters are pronounced and their individual pronunciations are not altered significantly by the letter contexts in which they are embedded.

The network of letter-to-phoneme connections for a single letter position exhibits characteristics that are dictated by the bi-alphabetic nature of the language. Each letter unit connects to its corresponding phoneme unit. No phoneme units are duplicated. For the phoneme units connected to the unambiguous letter units shared by the two alphabets, there is one letter unit per phoneme unit. For all other phoneme units, there are two letter unit connections per phoneme unit. That is, each unique Cyrillic, each unique Roman, and each ambiguous letter unit, connects to a phoneme unit which is connected to one other letter unit. This pattern of connections is repeated across letter positions.

With the activation of phoneme units by letter units, a two-way interactive process is initiated between the phoneme units and word units. Word units reflect the phonemic precision of the orthography. Each word unit represents a particular ordering of phoneme units together with their stress pattern. The access of word knowledge is seen to take place principally through the phoneme level to word level connections.

The foregoing concept of grapheme-phoneme rules is contrary to the view advanced in the literature on English word recognition and word pronunciation (e.g., Coltheart, 1978; Venezky, 1970). It rejects the idea that a rule expresses a regularity in which a single letter corresponds to a single phoneme. It promotes the idea that "rule" refers to any letter-to-phoneme correspondence in the language (cf. Rosson, 1985; Rumelhart & McClelland, 1982). In brief, rule is replaced by connection (see also Van Orden, 1987). In English, the number of letter/phoneme rules interpreted in the preceding sense would be very large; there would be many connections representing the many relations at the grain size of individual letters, and at coarser grains involving multiple letters. In Serbo-Croatian, by contrast, the number of rules-cum-connections is small. Besides number, English rules and the Serbo-Croatian rules contrast in mean strength, specifically, the mean level of activation of a letter unit required to activate a phoneme unit. In Serbo-Croatian the rules are stronger on average and the uniformity of strengths across the letter-to-phoneme connections is greater. The significance of these contrasts is highlighted in the comparative dependence of pronunciation in the two languages on word level information. Paraphrasing and generalizing Rosson (1985), because the rules of Serbo-Croatian are few and strong (on the average), word level information will be less critical; because the rules of English are many and weak (on the average) word level information will predominate (but see Van Orden, 1987, for an alternative view).

Let us summarize how the present data bear on the preceding account of word processing in Serbo-Croatian. A phonemically similar context activates almost all of the phoneme units of the following target. The shared and unshared phoneme units, in turn, will activate all of the word units that contain these phonemes in the same positions. On presentation of the target, this patterning of phoneme and word units, and the interaction among them, comprises the background against which the target's processing takes place.

The task of lexical decision is constrained primarily by the individual states and collective dynamics of the word units. When a context and target are phonemically similar, and the target is a highly familiar item, the processing of the context results in inhibition of the target's representation. This is because, in consensus with Grossberg's (1978) principle of self-modulation, the proportion of the net inhibition felt by a given unit (in a set of activated units) is scaled positively to the given unit's level of activity, which is scaled positively to the given unit's activation threshold, which is scaled positively to the familiarity of the word that the given unit represents. As a consequence, high familiarity targets are subject to high levels of inhibition during the processing of the context, with the result that phonemic similarity effects with high familiarity targets are negative, as evident in Experiments 3 and 4.

For low familiarity target words and for pseudowords, a phonemically similar context will activate most of the phoneme units to be used in processing the target. In neither case, however, will there be an exceptionally high level of activation in any individual word unit. Consequently, no individual word unit will be subject to high levels of inhibition. For a low familiarity word target, the low level of inhibition at the word unit level means that the target's processing can benefit from the pre-activation of the target's phoneme units and its word unit. Similarly, processing of a pseudoword target can benefit from the pre-activation of the target's phoneme unit. Nevertheless, if the lexical decision process is understandable ultimately as a signal detection problem, then decision speed and accuracy will vary inversely with the signal (activity level of target word unit)-to-noise (activity levels of other word units) ratio. Generally, this ratio would be smaller for pseudoword targets than for word targets. For pseudoword targets, therefore, the processing advantages of priming units below the word unit level could be outweighed, more or less, by post-lexical processing disadvantages. Differences in the lexical decision task between the phonemic similarity effects of low familiar and high familiar words were shown by the similarity by target familiarity interactions of Experiments 3 and 4. Differences in the lexical decision task between the phonemic similarity effects of pseudowords and words were shown in the similarity by lexicality interactions of Experiments 2 , 4, and 9.

The task of naming letter strings is constrained primarily by the individual states and collective dynamics of the phoneme units. A context that primes the phoneme units of a subsequent letter string should facilitate naming. Moreover, according to the model, it should do so equally for words and pseudowords. In isolation, naming a word is distinguished from naming a pseudoword by the contribution of information from accessed word units. When a word is presented, one word unit achieves dominance at the word unit level, reinforcing the states of the phoneme units. When a pseudoword is presented no word unit, on the average, achieves dominance, and the reinforcement of phoneme units is, therefore, much less pronounced. The lexicality variable, by definition, is manifest primarily at the word unit level—it is a matter of whether or not a word unit is fully activated in the processing of a letter string. The phonemic similarity variable operates, primarily, at the level of phoneme units—it is a matter of whether or not the target's phoneme units are activated. Consequently, in naming the effects of the one variable should add to the effects of the other variable. Facilitation by phonemic similarity, and an additive relation of phonemic similarity and lexicality, were found in Experiments 5, 6, and 7.

The significance of target familiarity in the phonemic similarity effect with lexical decision is understood in terms of the degree of inhibition generated at the word unit level. Inhibition, however, does not occur between adjacent levels, only activation, meaning that phonemes relevant to the processing of the target will not be inhibited even if the word unit relevant to the processing of the target is inhibited. For both high familiar and low familiar words, a phonemically similar context word will activate phoneme units relevant to naming the target. Downward activation from the word unit level may boost these phoneme units, more or less. Although phonemic similarity and target familiarity affected naming, Experiment 7 showed that they did not interact.

The articulatory units in the model are assumed to connect to output phoneme units and to word units. The former connection specifies the phonemic content of the pronunciation, the latter connection specifies the prosody of the pronunciation. A phonemically similar context whose accent differs from that of the target will specify appropriate aspects of the target's phonemic content but inappropriate accent. A phonemically similar context of the preceding type should affect naming; it should not, however, affect lexical decision. Experiments 8 and 9 confirmed these features of the model. Only naming was affected by accent incongruity. Accent incongruity appears to slow the assembling of a pronunciation more than the specification of phonemic content speeds up the assembling of a pronunciation.

In sum, the pattern of results obtained in the present series of experiments can be given a consistent interpretation on the assumption that phonological information in the processing of Serbo-Croatian words is computed automatically prior to lexical access. Both lexical access and the assembling of pronunciations depend on this prelexical phonology.

REFERENCES

Aaronson, D., & Ferres, S. (1983). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 700-725.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24, 89-106.

Baron, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology*, 25, 241-246.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 556-566.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 252-259.

Carello, C., Lukatela, G., & Turvey, M.T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition*, 16, 187-195.

Colombo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 226-234.

Coltheart, M. (1978), Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological coding in the lexical decision task. *Quarterly Journal of Experimental Psychology, 31*, 489-505.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). New York: Academic Press.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition, 6*, 391-402.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology, 33*, 325-350.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research, 45*, 55-72.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 288-298.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. Walker (Eds.), Sentence processing (pp. 27-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Forster, K. I. (1987). Form-priming with masked primes: The best match hypothesis. In M. Coltheart (Ed.), *Attention and Performance, XII* (pp. 128-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *Quarterly Journal of Experimental Psychology, 39A*, 211-251.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 361-379.

Frost, R., Bentin, S., & Katz, L. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256-281.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674-691.

Green, D. W., & Shallice, T. (1976). Direct visual access in reading for meaning. *Memory and Cognition, 4*, 753-758.

Grossberg, S. A. (1978) A theory of visual coding, memory, and development. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception*. New York: Wiley.

Hanson, V. L., & Fowler, C. A. (1987). Phonological coding in word reading: Evidence from hearing and deaf readers. *Memory & Cognition, 15*, 199-207.

Hawkins, H. L., Reicher, G., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 235-242.

Henderson, L. (1982). *Orthography and word recognition in reading*. New York: Academic Press.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: The encoding-bias hypothesis reconsidered. *Memory & Cognition, 8*, 115-123.

Humphreys, G. W., Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual role theory of reading. *The Behavioral and Brain Sciences, 8*, 689-739.

Humphreys, G. W., Evett, L. J., & Taylor, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition, 10*, 576-590.

Jordan, M. I. (1986). *Serial order in behavior: A parallel distributed processing approach* (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Studies.

Kay, J., & Marcell, A. J. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology, 33A*, 397-415.

Kolers, P. A. (1970). Three stages of reading. In H. Levin & J. P. Williams (Eds.), *Basic studies in reading* (pp.90-118). New York: Basic Books.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lukatela, G., Carello, C., Savić, M., & Turvey, M. T. (1986). Hemispheric asymmetries in phonological processing. *Neuropsychologica, 24*, 341-350.

Lukatela, G., Carello, C., & Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Language and Cognitive Processes, 2*, 1-17.

Lukatela, G., Feldman, L. B., Turvey, M. T., Carello, C., & Katz, L. (1989). Context effects in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 214-236.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition, 8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech, 21*, 142-165.

Lukatela, G., & Turvey, M. T. (1980). Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J.F. Kavanagh & R.L. Venezky (Eds.), *Orthography, reading, and dyslexia* (pp. 227-247). Baltimore, MD: University Park Press.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 237-254

Martin, R. C., & Jensen, C. R. (1988). Phonological priming in the lexical decision task: A failure to replicate. *Memory and Cognition, 16*, 505 -521.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88*, 357-407.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89*, 217-245.

Merikle, P. M, & Glick, M. I. (1976). Processing order in visual perception. *Quarterly Journal of Experimental Psychology, 28*, 17-26.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition, 2*, 309-321.

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition, 22*, 93-136.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review, 89*, 573-594.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonemic activation in silent word reading:

Evidence from backward masking. *Journal of Memory & Language, 27*, 59-70.

Rosson, M. B. (1985). The interaction of pronunciation rules and lexical representations in reading aloud. *Memory & Cognition, 13*, 90-98.

Rubenstein, H., Lewis., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645-657.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 1. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*, 60-94.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre-and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12*, 315-328.

Shulman, H. G., Hornak, R., & Sanders, E. (1978). The effect of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory & Cognition, 6*, 115-123.

Smith, F. (1971). *Understanding reading*. New York: Holt, Rinehart, & Winston.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112*, 1-36.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Orden, G. C. (1987). A ROSE is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*, 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 5*, 385-399.

West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition, 12*, 31-45.

FOOTNOTES

[*] *Memory & Cognition, 18*, 128-152 (1990).

[†] University of Belgrade.

[††] Also University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 74-84

Automatic and Prelexical Computation of Phonology in Visual Word Identification*

Georgije Lukatela† and M. T. Turvey††

Two experiments were conducted to evaluate the hypothesis that identification of a visually presented word involves phonological information that is activated prelexically and automatically. A backward masking paradigm was used in which a word target was followed by a pseudoword mask, followed in turn by a nonlinguistic pattern mask. The stimulus materials were Serbo-Croatian. The pseudoword mask could share all but one phoneme in common with the target, or none; moreover, it could be printed in the same alphabet as the target (for example, both stimuli printed in Cyrillic), or in the other alphabet (for example, target in Cyrillic, mask in Roman). Word targets were always in lower case, and pseudoword masks were always in upper case. It was assumed that where a mask shares phonological information with the target it can compensate for the interruption in processing by continuing the activation of the phoneme units activated by the target. Such an effect would be prelexical because the phoneme units activated by the mask would be those activated previously during the incomplete processing of the target. Both experiments, using different onset asynchronies among the stimuli, found significantly higher levels of target identification for homophonous masking than for nonhomophonous masking in agreement with previous studies of like kind using English materials. It was also shown that alphabet congruity affected the magnitude of the phonological effect in a direction that supported an hypothesis of inhibition of the letter processing units of one alphabet by the unique letters of the other alphabet. The results were discussed in terms of phonology's role in mediating lexical access in Serbo-Croatian and English, and in terms of a network model of visual word identification in Serbo-Croatian.

A long standing problem in the psychology of reading is whether or not the accessing of a written word's representation in the internal lexicon depends on the speech form of the word. Is the graphemic input recoded phonologically prior to lexical access, or does lexical access proceed in unmediated fashion? Research with Serbo-Croatian, which uses a phonetically precise orthography, tends to favor the view that the speech form of a word is activated prelexically and automatically in visual word recognition (e.g., Feldman & Turvey, 1983; Lukatela, Feldman, et al., 1989; Lukatela, Turvey, et al., 1989; Lukatela & Turvey, 1990). In the present article we seek further confirmation of the routine activation of phonology prior to lexical access.

Our previous research has used the methods of rapid lexical decision and, to a lesser degree, rapid naming. These methods rely on processing that is sufficiently complete to access the internal lexicon and may, therefore, be open to the criticism that the results obtained could be due to processes that arise subsequent to lexical access. This criticism would not apply with the same force to results obtained from visual masking situations. When a briefly exposed target word is followed closely in time by a pattern, the processing of the target word can be interrupted so that the information derived from the target is partial rather than full. If the temporal and intensity parameters of masking were such that the target word was identified rarely, then we would have to conclude that the performance level reflected the fact that

---

This research was supported in part by National Institute of Child Health and Human Development Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively.

the internal lexicon was accessed rarely. That is, the mask's effect was such that it frequently prevented processing from yielding enough information to access the target's lexical entry. Now, suppose that for the same physical parameters we found that a new masking stimulus, identical in visual complexity to the original mask, but related on some linguistic dimension to the target, resulted in a much higher level of target identification. Then we would have to assume that the rise in performance was brought about by the new mask contributing in some way to the processing of the target, specifically, to the processes needed for accessing the target's lexical representation.

The method employed in the present article, due to Naish (1980), follows the logic identified in the preceding paragraph. It entails the identification of briefly exposed target words under backward masking conditions with the following key features: The masks are (a) pseudowords, (b) phonologically related or unrelated to the targets, and (c) themselves followed by patterned stimuli to reduce their identification and, thereby, guessing strategies about target/mask relationships. The figural structure of the masks, and comparable intensities of the targets and masks, confine the masking effects on the targets to primarily central processes (Michaels & Turvey, 1979; Turvey, 1973). With respect to Naish's (1980) procedure, Perfetti et al. (1988) argued that if phonology is computed automatically, then phonological similarity between the mask and target will reduce the interruption of central processing normally induced by the mask. They reasoned that a phonologically similar mask will reinforce the phoneme processing units activated partially by the target. In contrast, a phonologically dissimilar mask will activate partially other phoneme units. If it is the case that word unit activation follows from phoneme unit activation, then a target preceding a phonologically similar mask will be identified better than a target preceding a phonologically dissimilar mask. The idea is that word units partially activated by a target will be activated further by a subsequent mask with common phonological properties. The outcomes of Naish's (1980) experiment and Perfetti et al.'s (1988) experiment (both using native speakers/readers of English and English language materials) were in agreement with this prediction.

A recently formulated model of Serbo-Croatian word recognition focuses on the interactions within and among the levels of letter processing units, phoneme processing units, and word processing units (Lukatela, Turvey, et al. 1989). It will be helpful to repeat the preceding masking argument in the context of this model.

According to the model, word identification is the product of a network of connected processing units. The connections among processing units for a single letter position honor the following assumptions. First, the letter units of the Cyrillic and Roman alphabets transcribing Serbo-Croatian constitute functionally distinct sets (Lukatela, Savić, Ognjenović, & Turvey, 1978). Common to the Roman and Cyrillic sets are the letter units corresponding to the letters E, A, O, J, K, M, and T. These letters are shared by the two alphabets. Each signifies the same phoneme in Roman and Cyrillic. In addition to this shared subset of letter units, the set of Cyrillic letter units consists of units corresponding to each of the letters that are unique to the Cyrillic alphabet, and units corresponding to each of the ambiguous letters. The upper case ambiguous letters are H, P, B, C. These letters signify different phonemes in the two alphabets. A similar structure is defined for the Roman letters. In addition to the shared subset of letter units, the set of Roman letter units consists of units corresponding to each of the unique Roman letters, and units corresponding to each of the ambiguous letters. A major idea is that at the level of letter units, each ambiguous letter is represented twice—once in the Cyrillic set and once in the Roman set.

The second assumption is that each letter unit connects to its corresponding phoneme unit. No phoneme units are duplicated. For the phoneme units connected to the shared unambiguous letter units, there is *one* letter unit connection per phoneme unit (e.g., A connects to /a/, E connects to /e/, and so on). For all other phoneme units, there are *two* letter unit connections per phoneme unit. That is, each unique Cyrillic, each unique Roman, and each ambiguous letter unit, connects to a phoneme unit which is connected to one other letter unit (e.g., Л and L both connect to /le/, Г and G both connect to /ge/, B and V both connect to /ve/, and so on). These letter unit-phoneme unit connections embody the letter-phoneme correspondence rules of the Serbo-Croatian language.

The third assumption is that, with the activation of phoneme units by letter units, a two-way interactive process is initiated between the phoneme unit and word unit levels. Word units in the proposed network reflect the phonemic precision of the orthography. Each word unit represents a particular ordering of phoneme units.

In consequence, in this model the accessing of word knowledge is seen to take place principally through the phoneme level to word level connections.

The preceding assumptions, as noted, are with respect to the organization of connections between levels for the processing of a single letter position. The letter units-phoneme units organization just described repeats for each letter position (cf. McClelland & Rumelhart, 1981). That is, if the model is addressing the processing of a five letter word, then there will be five sets of letter units-phoneme units connections, one for each letter position. The most important assumption of the present model of single word processing is made with respect to the relations between these letter-phoneme organizations across letter positions. It is assumed that, across letter positions, there are multiple inhibitory connections in both directions between the unique letter units of one alphabet and the unique and ambiguous letter units of the other alphabet. When a unique Cyrillic (Roman) letter unit is activated by a stimulus letter in one position, then the activity in all Roman (Cyrillic) letter units in the other letter positions is reduced.

Let us now return to Naish's (1980) masking paradigm. A target stimulus written with unique letters can be expected to induce inhibition of the letter units of the other alphabet, to excite the phoneme units specified by its orthography, and to excite word units containing these phonemes. The degree to which the target does these things will depend on how much time elapses before the mask arrives. A pseudoword masking stimulus that is transcribed in the same alphabet as the target, and which shares all phonemes but one in common with the target, will reinforce the inhibitory activity, most of the phoneme unit activity, and most of the word unit activity, initiated by the target. A pseudoword mask in the other alphabet, and possessing no phonemes in common with the target, will reverse the inhibitory activity among letter units, and excite different phoneme units and, thereby, different word units. A phonologically similar, same alphabet mask, should yield higher target identification than a phonologically dissimilar, other alphabet mask but, more generally speaking, target word identification should be higher under phonologically similar masking conditions regardless of the alphabetic relation between the stimuli.

The model of Serbo-Croatian word processing sketched above suggests a further prediction that we will refer to as the alphabet priming hypothesis. Let us consider more closely the inhibitory process at the letter unit level. When target and mask differ in alphabet, the letter units for the mask will be activated more slowly due to their previous inhibition by target letter units. Consequently, the mask will abrogate processing at the phoneme unit level more slowly. This retardation in mask processing will have two related effects in the context of Naish's (1980) paradigm. One effect is that the target's phoneme units will have more time to activate uninterruptedly the appropriate word unit candidates. The other effect is that the target's phoneme units will receive less of a boost from a mask with common phonemic constituents. These effects of an alphabetic difference between target and mask should show up in the data as follows: (a) target identification will be higher when followed by a phonologically dissimilar mask in the other alphabet than when followed by a phonologically dissimilar mask in the same alphabet, and (b) target identification will be lower when followed by a phonologically similar mask in the other alphabet than when followed by a phonologically similar mask in the same alphabet.

In the two experiments reported, we manipulate the relations between target and mask stimuli in terms of phonology and alphabet, and we assess the effects of these relations at onset asynchronies of 40 ms (target to mask) and 40 ms (mask to pattern) (Experiment 1) and 20 ms and 20 ms (Experiment 2). In both experiments we look for an effect of phonological similarity and the particular interaction between phonological similarity and alphabet identified in (a) and (b) above. Our expectation is that both experiments will show a phonological similarity effect, but that the second experiment with the shorter time scale will be more likely to reveal the interaction.

EXPERIMENT 1

We take advantage of certain features of written Serbo-Croatian to provide a strong test of prelexical computation of phonology in Naish's backward masking paradigm. First, visual similarity between a target and a phonologically similar mask is ruled out by presenting the two stimuli in different cases and different alphabets. As noted above, Serbo-Croatian has two writing systems, Roman and Cyrillic, that use mostly different letters to represent the phonemes of the language (see Lukatela and Turvey, 1980). Consequently, by combining the two alphabets and the two cases it is possible to transcribe a target word and its phonologically similar mask such that there are often no shared visual forms and at worst only one shared visual form. Second, because of the phonemic precision of the orthography, a word and its pseudoword rhyme will have exactly the same number of letters. This feature protects against the possibility of non-superimposition of target (e.g., nought) and mask (e.g., nort) which opens the door for visual explanations of the phonological effect (Naish, 1980). Third, in English experiments phonemic similarity involving pseudowords cannot always be defined unequivocally. English language experimenters have to use an operational definition, namely, whether or not the pseudowords in question produce a pseudohomophone effect in the lexical decision task (Humphreys, Evett, & Taylor, 1982). The present experiment uses word-pseudoword target-mask pairs for which the phonemic similarity relation is unambiguous. This feature follows from the facts that the Serbo-Croatian writing system is phonemically precise and that all letters in a word are pronounced. Thus, a word context that differs from a pseudoword target in one letter is phonemically identical to the target with respect to all letter positions but one.

Method

Subjects

Sixteen high-school juniors from the Fourth Belgrade Gymnasium were paid for their participation in the experiment They were assigned randomly to one of four counterbalancing groups according to their appearance at the laboratory.

Materials

Fifty-six target words were selected from the middle familiarity range of a word list established by an extended series of familiarity judgement tests (cf. Gernsbacher, 1984). These tests were run with other high-school juniors who did not participate in the experiment. All selected words were Serbo-Croatian masculine nouns in nominative singular, and they had the CVCVC structure. Twenty-eight target words were written in the Cyrillic lower case (for example, "шешир" means "the hat"), and twenty-eight in the Roman lower case (for example, "labud" means "the swan"). All target words were paired with each of four mask types: (i) the same alphabet, homophonic, (ii) the same alphabet, nonhomophonic control, (iii) the other alphabet, homophonic, and (iv) the other alphabet, nonhomophonic control. All mask types were pseudowords of the CVCVC structure, that is, each mask was of the same length and structure as the target word. In Same Alphabet Homophonic pairs both the target and mask were written either in the Cyrillic or in the Roman alphabet. The masking letter string was generated by replacing the final consonant of the target word. By this procedure, in each Same Alphabet Homophonic pair, and in each Other Alphabet Homophonic pair, only the final phoneme was not shared by the target and mask. Contrary to this, in all Nonhomophonic pairs no phonemes/graphemes were shared by the target and mask. For example, as illustrated in Table 1, for the target "labud," the four mask types were LABUG (same alphabet, homophonic), ZIVOK (same alphabet, nonhomophonic control), ЛАБУГ (other alphabet, homophonic), and ЖИВОК (other alphabet, nonhomophonic control).

Table 1. *Sample stimuli used in experiments.*

Target word	Masking pseudoword	
	Phonologically similar	Phonologically dissimilar
Same alphabet (Cyrillic-Cyrillic):		
шешир	ШЕШИЛ	РУЧАЗ
ручак	РУЧАЗ	ШЕШИЛ
Same alphabet (Roman-Roman):		
labud	LABUG	ZIVOK
zivot	ZIVOK	LABUG
Other alphabet (Cyrillic-Roman):		
шешир	SESIL	RUCAZ
ручак	RUCAZ	SESIL
Other alphabet (Roman-Cyrillic):		
labud	ЛАБУГ	ЖИВОК
zivot	ЖИВОК	ЛАБУГ

In addition to the experimental 56 words, another set of 72 five-letter words was selected also from the mid-familiarity range. This other set was used to prepare three groups of foils designed to protect against response biases that might be induced by repeating fixed relations between the experimental stimuli. The first group consisted of

20 Identity Word-Word pairs in which the same word appeared both as the target and as the mask (e.g., dolar - DOLAR). The second group contained 20 Initial Different Word-Nonword pairs in which the initial consonant was not shared by the target and mask (e.g., zapad - FAPAD). The third group of foils had 32 Blank Field-Word pairs in which a blank field preceded a five letter word in the mask position (e.g, - KAZAN).

Design

A given subject never encountered a word or pseudoword in any of the pairs more than once but every subject saw every type of pair. These conditions were met with four counterbalancing groups. In total, each subject saw 56 experimental plus 72 filler pairs ordered randomly.

Procedure

The subject was comfortably seated before the cathode ray tube of an Apple IIE computer in a dimly lit room. Target words were displayed in lower case and pseudoword masks in uppercase in a dark on light format. The target and pseudoword mask had nominal exposures of 40 ms each. These exposure durations are "nominal" rather than exact, because display changes in reality occurred within the standard 16-ms scan rate of the Apple IIE monitor. This means that all actual durations of the nominally 40-msec exposure varied in a random manner with a uniform probability between 40 and 56 ms, whereby the statistical mean exposure was 48 ms.

Each trial was initiated by a subject pressing the space bar on the keyboard. Then, a fixation point appeared in the center of the screen and the subject heard a brief warning signal after which a target (or a blank field) appeared for the nominal 40-msec duration replacing the fixation point. The target was followed (with a zero interstimulus interval) by a pseudoword mask, which replaced the target. The pseudoword mask was presented for 40 ms and, then, it was substituted by a row of six hashes which remained on the screen until the next trial. The subject was instructed to write down on a prepared form the word he or she believed to have seen during the brief presentation of stimuli. The subject was encouraged to write, in default, just a single letter.

Results and Discussion

The mean percent correct identifications of experimental targets and their standard deviations are shown in Table 2. Percentages of experimental targets identified correctly were analysed twice, once with subjects and once with items as a random effect. The *F* tests were then combined using Clark's (1973) min *F* ′ procedure. A 2 (Mask Type, phonologically similar vs. dissimilar) × 2 (Alphabet, same vs. other) analysis of variance was performed. Only the main effect of Mask Type was significant, min $F'(1, 59) = 43.41$, $p < 001$, with the phonologically similar masks eliciting better identification of targets (59%) than the phonologically dissimilar masks (27%). The main effect of Alphabet was not significant, min $F'(1, 62) = 1.71$, $p < .2$, nor was the interaction between Mask Type and Alphabet, min $F'(1, 63) = 2{,}91$, $p < .1$.

Table 2. *Mean percent correct identification M (in %) of the test targets, and standard deviations SD (in %) in Experiment 1 as a function of phonological similarity of target and mask, and alphabet of mask relative to that of target.*

Alphabet	Phonologically similar		Phonologically dissimilar	
	M	SD	M	SD
same	66.37	15.76	27.26	17.44
other	52.80	19.29	28.33	20.61

The results are in agreement with the primary hypothesis identified in the introduction, and confirm the previous observations of Naish (1980) and Perfetti et al. (1988): Masking was reduced under phonologically similar pseudoword masks. Naish (1980) had used onset asynchronies of 30 ms and 20 ms. Perfetti et al. (1988) used three different pairs of onset asynchronies. They found the strongest effects of phonological similarity with the pair of shortest onset asynchronies tested, namely, 30 ms and 25 ms. Naish (1980) found a 14.5% effect, Perfetti et al. (1988) found a 22% effect (approximately). In the present experiment, the phonological similarity effect was 32%.

As noted above, we follow Perfetti et al. (1988) in assuming that this retroactive effect of a phonologically similar mask is on a process of identification that is incomplete at mask onset. Because the mask shares phonological information with the target, it compensates, in part, for the processing interruption it brought about by refreshing the phoneme units activated by the target and, thereby, the word units consistent with those phoneme units. The effect is prelexical because the phoneme units affected by the mask are those activated during the incomplete processing of the target.

With respect to the second hypothesis identified in the introduction, we preface its evaluation by considering in more detail what is at issue in the manipulation of alphabet. When a pseudoword mask in upper case and a word target in lower case are written in the same alphabet, is the condition one of alphabet congruity or graphemic similarity? In the processing model laid out in the introduction, alphabet is a variable of significance at the letter unit level of processing. At the letter unit level the two alphabets are distinguished within the network, with inhibitory connections defined reciprocally between the letter units of the Roman alphabet and the letter units of the Cyrillic alphabet. Activation of a unique letter processing unit of one alphabet induces inhibition in the set of letter processing units of the other alphabet. Alphabet congruity of target and mask means that the mask will activate noninhibited letter units; alphabet incongruity means that the mask will activate inhibited letter units.

Graphemic similarity can be understood in two ways depending on the sense of the term grapheme. Henderson (1985), in a critical review, identifies two senses in which 'grapheme' can be used. In Sense 1, the grapheme is the minimal contrastive unit in a given writing system. In Sense 2, a grapheme comprises a letter or letters that specifies a single phoneme in speech. From the point of view of Sense 1, a phonemically similar target and mask that differ in alphabet are graphemically dissimilar; from the point of view of Sense 2, a phonemically similar target and mask that differ in alphabet are graphemically similar. In either sense, an hypothesis of graphemic similarity of target and mask means that the mask will activate the same graphemic processing units as the target; graphemic dissimilarity means that the mask will activate different graphemic processing units.

Let us now look at the predictions from the two senses of grapheme, beginning with Sense 1. If word units are activated by graphemic information, then we should expect that targets followed by masks that are phonemically similar and in the same alphabet (that is, graphemically similar) should be identified better than targets followed by masks that are phonemically similar and in the other alphabet (that is, graphemically dissimilar). We should also expect that for the two conditions of phonemically dissimilar masks, that is, same alphabet and other alphabet, there should be no difference in target identification; both conditions lack graphemic similarity, by definition.

Let us now look at the prediction from Sense 2. When two stimuli are phonemically similar they are graphemically similar, regardless of the alphabet relation between them. Consequently, targets followed by masks that are phonemically similar and in the same alphabet, and targets followed by masks that are phonemically similar and in the other alphabet, should be identified with the same accuracy. Similarly, for the two conditions of phonemically dissimilar masks, there should be no difference in target identification; both conditions lack graphemic similarity, by definition.

In sum, with respect to the homophonous mask conditions of the present experiment, the Sense 1 and Sense 2 interpretations of graphemic similarity make different predictions; with respect to the nonhomophonous mask conditions, they make the same prediction. With same alphabet/phonemically similar masks percent identification was 66.1%; with other alphabet/phonemically similar masks percent identification was 52.2%. The two conditions of nonhomophonous masks yielded 26.3% and 28.1%. Although this pattern comports with the Sense 1 hypothesis of graphemic similarity, there was no significant interaction between alphabet match and phonological similarity, so it must be concluded that the 66.1% to 52.2% was not a significant difference.

With respect to the question of whether or not alphabet per se was a factor, rather than graphemic similarity in either Sense 1 or Sense 2, it will be recalled from the introduction that the significance of alphabet would be shown by way of a particular kind of interaction between alphabet match and phonological similarity. Namely, that for the homophonous mask conditions, same alphabet should be superior to other alphabet *but* for the nonhomophonous mask conditions, other alphabet should be superior to same alphabet. The alphabet prediction for homophonous masks agrees with the Sense 1 graphemic prediction. The alphabet prediction for nonhomophonous masks agrees with neither the Sense 1 nor Sense 2 predictions. We reiterate that the statistical analysis of the alphabet match by phonological similarity interaction did not find significance. On the basis of this first experiment, therefore, we cannot confirm the secondary hypothesis based on alphabet inhibition nor, relatedly, can we distinguish between the graphemic and alphabet interpretations of the alphabet manipulation.

EXPERIMENT 2

The second experiment repeats the design of the first but with briefer onset asynchronies among the three stimuli (target word, pseudoword mask, pattern mask). Our expectation is that the hypothesized alphabet match by phonological similarity interaction expected from the alphabet priming hypothesis is more pronounced under some parameters of stimulus presentation than others. The parameters in question are likely to be those at which processing time is so limited that the retarding of mask processing due to alphabet inhibition is not compensated by the mask's contribution to phoneme unit activation. Our guess is that such parameters comprise an extremely brief onset asynchrony between the word and the following pseudoword, and an extremely brief onset asynchrony between the pseudoword and the following pattern.

In the second experiment the onset asynchronies were half those of the first experiment. It was hoped that the data obtained at these onset asynchronies would distinguish among the two graphemic similarity hypotheses and the alphabet priming hypothesis. The predictions of the three hypotheses are summarized in Table 3. With respect to the major hypothesis of a phonological similarity effect, it was expected that the second experiment would replicate the results of the first.

Table 3. *Predictions of the relative mean correct target identification for the different mask conditions under three different hypotheses.*

Hypothesis	Prediction		
Alphabetical Bias	SH > SNH	SH > OH	SNH < ONH
Sense 1 Graphemic	SH > SNH	SH > OH	SNH = ONH
Sense 2 Graphemic	SH > SNH	SH = OH	SNH = ONH

SH = Same Alphabet Homophonic
OH = Other Alphabet Homophonic
SNH = Same Alphabet Nonhomophonic
ONH = Other Alphabet Nonhomophonic

Method

Subjects

Twenty high-school juniors from the Fourth Belgrade Gymnasium were paid for participation in the experiment. They were assigned randomly to one of four groups according to appearance at the laboratory. None of them participated in Experiment 1.

Materials and Design

These were the same as for Experiment 1.

Procedure

The procedure of Experiment 1 was used, except that the onsets of the word target and the pseudoword mask, and the onsets of the pseudoword mask and the pattern mask, were separated by a nominal 20 ms.

Results and Discussion

The mean percent correct identifications of experimental targets and their standard deviations are shown in Table 4. A 2 (Mask Type, phonologically similar vs. dissimilar) × 2 (Alphabet, same vs. other) analysis of variance proved that the main effect of Mask Type was significant, min F' (1, 45) = 15.49, $p < .001$, with the phonologically similar mask type eliciting better identification of targets (28%) than the phonologically dissimilar mask type (10%). The main effect of Alphabet was not significant, min $F' < 1$, but the interaction between Mask Type and Alphabet reached significance, min F' (1, 68) = 4.18, $p < .05$. Protected t-tests showed that between Same Alphabet Homophonic pairs and Other Alphabet Homophonic pairs there was no statistical difference, min $F' < 1$. For Nonhomophonic pairs, however, the analysis revealed a significant difference between Same Alphabet and Other Alphabet condition, min F' (1, 74) = 3.85, $p < .05$. Correct identification of the target words occurred more frequently in the Other Alphabet condition (13.6%) than in the Same Alphabet condition (7.1%).

Table 4. *Mean percent correct target identification M (in %) and standard deviation SD (in %) in Experiment 2 as a function of phonological similarity of target and mask, and alphabet of mask relative to that of target.*

Alphabet	Phonologically similar		Phonologically dissimilar	
	M	SD	M	SD
same	32.14	20.66	7.14	8.03
other	24.29	16.45	13.57	7.29

The present experiment reinforced the phonological similarity effect in backward masking observed in Experiment 1, and in the experiments of Naish (1980) and Perfetti et al.

(1988). In addition it confirmed the secondary hypothesis identified in the introduction, a hypothesis that was derived specifically from the model of Serbo-Croatian word processing (Lukatela, Turvey et al., 1989): The effect of phonological similarity was less when the word target and pseudoword mask were printed in different alphabets than when they were printed in the same alphabet. More specifically, in nonhomophonous mask conditions, pseudoword masks printed in the other alphabet produced superior target identification than pseudoword masks printed in the same alphabet. This outcome favors the view that the manipulation of alphabet in the two experiments was not a manipulation of graphemic similarity in either Sense 1 or Sense 2 (see Table 3). It adds to a number of other results with Serbo-Croatian materials that suggest a rapidly established, but short-lived, inhibition of letter processing units in terms of alphabet affiliation: The presentation of unique Roman letters induces inhibition of the Cyrillic letter-to-phoneme correspondencies; the presentation of unique Cyrillic letters induces inhibition of the Roman letter-to-phoneme correspondencies (Lukatela, Feldman et al., 1989; Lukatela, Turvey, et al., 1989).

GENERAL DISCUSSION

The present results can be viewed as a demonstration that information about the phonology of a printed word becomes available (a) prefatory to lexical access, and (b) automatically. A higher level of target identification under homophonic masking relative to nonhomophonic masking can be said to reflect the continued activation, under homophonic masking, of phoneme units activated initially in the incomplete processing of the target. With respect to the autonomy claim, the design of the experiments did not encourage phonemic recoding—asking only that subjects write down what they saw—nor did it provide opportunity for task-specific strategic processes—such as guessing a target on the basis of its homophony with the mask. Across stimuli the homophonous masks were paired equally often with phonemically dissimilar targets. Moreover, the pseudoword masks were themselves masked.

There is some consensus that English word identification usually proceeds without recourse to phonology, and that the activation of phonological information is optional (Humphreys & Evett, 1985). Evidence occasionally obtained in lexical decision experiments for the involvement of phonological properties is attributed to postlexical processes. The earlier results of Naish (1980), and the recent results of Perfetti et al. (1988), and van Orden (1987; van Orden, Johnston, & Hale,1988) do not concur with the consensus view, however. In van Orden's experiments, the subject had to decide whether or not a printed word was a member of a given semantic category. Categorization errors were observed frequently when homophones of category instances were presented, for example, *rows* is homophonic with *rose*, and subjects tended to categorize *rows* as a flower more so than its spelling control *robs*. Further, false positive errors to nonword homophone foils (e.g., *sute* for *an article of clothing*) exceeded false positive errors to nonhomophonic nonword spelling controls; the phonological characteristics of the nonword foils proved to be critical. Van Orden and his colleagues interpreted these results as demonstrating a strong prelexical phonological component in visual word identification. Given (a) the parallelism between the results in the English language studies of Naish and Perfetti et al., and the present results with Serbo-Croatian, and (b) the converging experimental evidence that Serbo-Croatian word identification is phonologically mediated, it seems prudent to suggest that for both orthographies lexical access depends on prelexical phonological information. Phonological mediation may be true for English and for Serbo-Croatian, but the evidence is, apparently, harder to come by in experimental studies with the English orthography.

If visual word processing in both Serbo-Croatian and English is phonologically mediated, then the data differences between Serbo-Croatian and English studies must reflect differences that arise within the mechanism of phonological coding, and not differences that arise from contrasting ways (visual vs. phonological) of accessing word information. Van Orden (1987) suggests that the mechanism that connects orthographic and phonological features is sensitive to the covariation between these two kinds of features across many words. The associative weights established between orthographic and phonological features would be determined by both the frequency and the consistency of covariation. Serbo-Croatian and English differ in the covariation of orthographic and phonological features, a difference that is captured by the contrast of shallow and deep orthography (Lukatela & Turvey, 1980). A difference in covariation (in frequencies and consistencies of relations) between the languages would induce a difference in the mechanism of prelexical phonology.

To date, the evidence for prelexical phonology in Serbo-Croatian visual word processing has been derived from two major strategies. One strategy exploits the special phonological ambiguity that arises in written Serbo-Croatian by virtue of the two, partially overlapping alphabets used to transcribe the language. The other strategy is based in the class of priming effects that are produced when a target word is preceded by a word similar to the target on one or more dimensions. Let us consider first the strategy based on the fact that Serbo-Croatian has two writing systems, Roman and Cyrillic. As noted above, these two alphabets share a number of letters, some of which designate one phoneme in one alphabet and another phoneme in the other alphabet (Lukatela & Turvey, 1980). There are written words in Serbo-Croatian that are composed of only shared letters. Some of these words contain phonemically ambiguous letters. Take BETAP as an example. Read strictly through the letter-to-sound correspondencies of the Cyrillic alphabet, this letter string is pronounced /vetar/ and is a high frequency noun meaning "wind". Read strictly through the letter-to-sound correspondencies of the Roman alphabet, BETAP is pronounced /betap/, a pseudoword. Read with a mixture of the two sets of correspondencies, Cyrillic and Roman, leads to the pronunciations /vetap/ and /betar/, which are also pseudowords. The word meaning "wind" is transcribed in the Roman alphabet as VETAR. This letter string supports only a single reading, /vetar/. No other readings are possible. VETAR, unlike its Cyrillic mate BETAP, is phonologically unambiguous. In sum, the primary distinction between a BETAP-type word and a VETAR-type word lies in the greater number of phonological descriptions supportable by BETAP's letter structure.

In the lexical decision, the acceptance latencies for BETAP, and for words like it, are considerably longer than the acceptance latencies for VETAR, and for words like it, even though BETAP and VETAR are equal in frequency, syllabic structure, number of letters, and meaning. This contrast defines what is referred to as the phonological ambiguity effect. It is argued that accepting BETAP as a word is made more difficult by the fact that it activates more phoneme units than VETAR which, in turn, activate partially many more competing word units than VETAR (Lukatela, Turvey, et al., 1989). Importantly, the phonological ambiguity effect satisfies a major criterion advanced by Coltheart et al. (1977) and van Orden (1987) for the demonstration of phonological mediation in lexical access, namely, that phonological effects be demonstrated empirically in *positive* lexical decisions. A summary of the varied empirical phenomena associated with the phonological ambiguity effect that implicate an automatic computation of phonology are to be found in Lukatela and Turvey (1990).

The other major strategy that we have used to address prelexical phonology is that of evaluating the effects on rapid naming and lexical decision of phonological similarity between a target word and the word context that precedes it. Hillinger (1980) found that, in lexical decision, word sequences such as EIGHT-MATE resulted in shorter decision times for MATE than corresponding control sequences, suggesting that MATE benefitted from the phonological information activated by EIGHT. This result of a phonological effect independent of visual similarity has been demonstrated in several naming and lexical decision experiments with Serbo-Croatian, taking advantage of the two alphabets to produce homophonous pairs without visual similarity and without Sense 1 graphemic similarity (Lukatela & Turvey, 1990). Using a visual masking procedure borrowed from Humphreys, Evett, and Taylor (1982), Lukatela and Turvey (Experiment 10, 1990) showed that masked pseudowords reduced lexical decision times on immediately subsequent homophonous target words. The failure of Humphreys et al. (1982) to obtain this result led them to conclude against automatic activation of prelexical phonology. In view of the positive outcome of the Lukatela and Turvey (1990) experiment, and the parallel results of the present experiments and those of Naish (1980) and Perfetti et al. (1988), the conclusion of Humphreys et al. (1982) should be reconsidered.

As observed in the introduction, the data from experiments on word identification in Serbo-Croatian have motivated a network model in which phoneme processing units mediate between letter processing units and word processing units. The present results demonstrating a phonological effect in backward masking are consistent with this model. Also consistent with the model is the demonstration in Experiment 2 of an interaction between alphabet match and phonological similarity, where alphabet mismatch yields superior identification, in the absence of phonemic similarity, than alphabet match. If a letter string containing letters unique to one of the two alphabets is presented, then letter units of the other alphabet are inhibited. An immediately following letter string that shares no graphemes/phonemes in common, will interrupt the processing of the first string. How much it does so depends, however, on the alphabet in which it is printed. If the second stimulus is printed in the other alphabet, then the letter units needed to process it will be inhibited, and more time will be needed for the second stimulus to overwrite the contents of the letter unit level. Consequently, the first stimulus will be processed further without interruption than if it had been followed by a letter string in the same alphabet.

Before concluding, an alternative view of the homophony effect needs to be addressed. Namely, the possibility that it was due to post-lexical processes. If the pseudoword mask was available to awareness, then information about it could have supplemented the information about the target at later stages rather than an earlier, prelexical stage. One prediction of this view is that at greater onset asynchronies between the pseudoword and the aftercoming pattern mask, the pseudoword should become more available to awareness and, consequently, the effect of homophony should increase. Although the onset asynchrony in the second experiment was half that in the first experiment, the significance level of the effect of homophony was the same ($p < 0.001$). Relatedly, in Perfetti et al.'s (1988) experiments, the interaction of homophony and onset asynchrony was not significant. These observations are counter to the notion that the main effect of homophony is linked to the pseudoword mask's availableness in awareness. A further counter argument follows from the alphabet effect of Experiment 2. Under nonhomophonic conditions, in which mask and target shared neither graphemes nor phonemes, pseudoword masks produced reliable influences on the correct identification of target words simply as a function of alphabet congruity. The absence of any true linguistic relation between the target and mask would seem to preclude any lexical or postlexical interpretation of the influence of alphabet; and the particular form of the influence makes sense only in terms of a prelexical network with specialized inhibitory connections across alphabets of the kind described above. Given that one and the same temporal and intensity conditions gave rise to both the homophony and alphabet effects, it seems reasonable to conclude, therefore, that both are prelexical. The preceding observations, though counter to the post-lexical hypothesis, cannot be taken as conclusive evidence against it. Future experiments will need to include more direct evaluations of this alternative interpretation.

In summarizing, we return the discussion to the model of Serbo-Croatian word recognition. The particular organization of the units in the processing network proposed by Lukatela, Turvey, et al. (1989) expresses the peculiarities of the Serbo-Croatian orthography—its phonetic precision and its bialphabetism. These peculiarities, embodied in the pattern of connections, lead to certain predictions about word identification under backward masking: Homophonous masks should aid target identification, and the contribution of mask homophony to target identification should vary with the alphabet relation between the stimuli. Both predictions were upheld. The implication of confirming the major prediction is that word identification is mediated by phonological information. The fact that the identical outcome was obtained in the present experiments and previous experiments of the same kind in English, suggests that word identification in these different orthographies proceeds similarly by reference to phonology. The implication of confirming the secondary prediction is that a general theory of word identification cannot ignore possible variations in the functional architecture of lower level mechanisms (for example, letter units and phoneme units) due to variations in orthography (Turvey, Feldman, & Lukatela, 1984).

REFERENCES

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 4*, 674-690.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535 -555). New York: Academic Press.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256-281.

Henderson, L. (1985). On the use of the term 'grapheme.' *Language and Cognitive Processes, 1*, 135-148.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: The encoding-bias hypothesis reconsidered. *Memory & Cognition, 8*, 115-123.

Humphreys, G. W., Evett, L. J., & Taylor, D. E. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition, 10*, 576-590.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. (1978). On the relation between processing the Roman and the Cyrillic alphabets: A preliminary analysis with bialphabetical readers. *Language and Speech, 21*, 113-141.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech, 21*,142-165.

Lukatela, G., & Turvey, M. T. (1980). Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. F. Kavanagh & R. L. Venezsky (Eds.), *Orthography, reading, and dyslexia* (pp. 227-247). Baltimore, MD: University Park Press.

Lukatela, G., & Turvey, M. T. (1990). Phonemic similarity effects and prelexical phonology. *Memory and Cognition, 18*, 128-152.

Lukatela, G., Feldman, L. B., Turvey, M. T., Carello, C., & Katz, L. (1989). Context effects in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 214-236.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetical word perception. *Journal of Memory and Language, 28*, 237-254.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 357-407.

Michaels, C. F., & Turvey, M. T. (1979). Central sources of visual masking: Indexing structures supporting seeing at a single brief glance. *Psychological Research, 4*,1-61.

Naish, P. (1980). The effects of graphemic and phonemic similarity between targets and masks in a backward visual masking paradigm. *Quarterly Journal of Experimental Psychology, 32*, 57-68.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59-70.

Turvey, M. T. (1973). On peripheral and central processes in vision: An information processing analysis with patterned stimuli. *Psychological Review, 80*, 1-52.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Orden, G. C. (1987). A ROSE is a ROSE: Spelling, sound and reading. *Memory and Cognition, 15*, 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

FOOTNOTES

[*] *European Journal of Cognitive Psychology*, in press.

[†] University of Belgrade.

[††] Also University of Connecticut.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 85-100

Phonological Ambiguity and Lexical Ambiguity: Effects on Visual and Auditory Word Recognition*

Ram Frost,† Laurie B. Feldman,†† and Leonard Katz†††

The present study was concerned with the relative effects of phonological ambiguity and lexical ambiguity on word recognition. The study exploited a special property of Serbo-Croatian. Specifically, readers of that language are familiar with two alphabets which share a subset of letters that have two different pronunciations, one in each alphabet. The stimuli in this study used letter strings that could be pronounced differently in each alphabet. Some of these strings had two different meanings, one for each pronunciation; for others, only one of the two pronunciations was meaningful. In three experiments, subjects were presented with an ambiguous printed word and, simultaneously, with a spoken word and were required to determine if they matched. The spoken words were presented clearly or degraded by noise. The speed of matching the spoken words to phonologically ambiguous letter strings was measured relative to their phonologically unequivocal controls. The results indicated that phonological ambiguity slowed stimulus matching. Moreover, phonological ambiguity had a greater effect when the phonologically ambiguous form represented two meaningful words. In neither simultaneous nor delayed presentation did auditory degradation interact with phonological or lexical ambiguity, suggesting that perception of the spoken word was independent from perception of the printed word.

In the past decade, the problem of lexical ambiguity has attracted the attention of many researchers concerned with models of printed word recognition. Since lexical ambiguity exists in many languages, an examination of the process of disambiguation may clarify the relation between orthography and phonology in the mental lexicon.

Ambiguity can exist in the relationships between the orthographic and phonological forms of a word, or between the phonological form and its semantic representation. The printed word "bank," for example, has one pronunciation (and therefore an unambiguous phonological representation), but two meanings. In contrast to"bank," the printed word "wind" has two different phonological representations (/wind/ and /waynd/) each of which has a different semantic meaning. If we consider a word to be, first of all, a phonological structure, and the word's phonological representation to be its lexical entry, then the printed form "wind" has two lexical entries, each of which is semantically differentiated. Resolving the phonological ambiguity necessarily designates one specific lexical entry which, in turn, automatically disambiguates the semantic representation (and vice versa). The upper portion of Figure 1 diagrams these relationships for English.

Effects of semantic ambiguity on visual word recognition have been demonstrated in several studies (for a review see Simpson, 1984). Rubenstein, Lewis, and Rubenstein (1971) showed that lexical decisions are faster for homographs like "bank," than for nonhomographic words. The explanation for this effect was that words with multiple meanings have multiple lexical entries, and therefore the probability of encountering one of them is greater than the probability of detecting

---

This work was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. We are indebted to Georgije Lukatela, Aleksandar Kostić, and the other members of the Laboratory for Experimental Psychology at the University of Belgrade, for their help, and to Shlomo Bentin, Keith Rayner, Charles Perfetti, and two anonymous reviewers, for their criticism on earlier drafts of this paper.

a single entry (see also Jastrembsky, 1981). Effects of phonological ambiguity were demonstrated in the naming task which, by its nature, requires a fully specified phonological form. Homographs like "wind" which are phonologically ambiguous, take longer to name than homographs with a single pronunciation (Kroll & Schweickert, 1978).

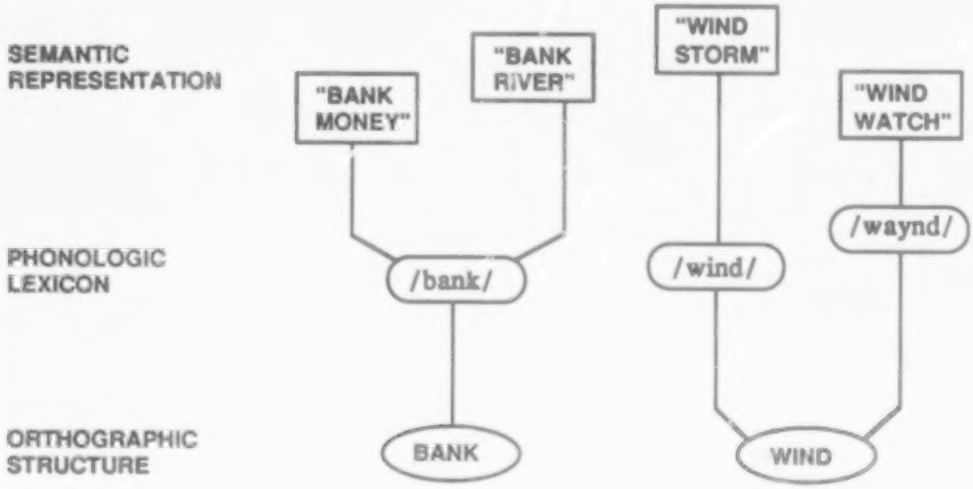
An additional source of phonological ambiguity can be demonstrated in languages with two distinct and equivalent orthographies like Serbo-Croatian. In Serbo-Croatian, the Roman and the Cyrillic alphabets are both taught to all elementary school children, and used interchangeably by the skilled reader. Most characters in the two alphabets are unique to one alphabet or the other, but there are some characters that occur in both. Of those, some receive the same phonemic interpretation regardless of alphabet (COMMON letters), but others receive a different interpretation in each alphabet (AMBIGUOUS letters). Letter strings that include unique letters can be pronounced in only one alphabet. Similarly, letter strings composed exclusively of common letters can be pronounced in the same manner in both alphabets. In contrast, strings that contain only AMBIGUOUS and COMMON letters are PHONOLOGICALLY BIVALENT. They can be pronounced in one way by treating the characters as Roman letters, and in a distinctly different way by treating them as Cyrillic letters. For example, the letter string "POTOP" can be pronounced as /potop/ if the ambiguous character *P* is interpreted according to its Roman pronunciation. By contrast, if the letter string is taken as a Cyrillic spelling, the grapheme *P* receives the pronunciation /r/ and the string must be pronounced /rotor/ (the characters O and T are common, and have the same pronunciation in the two alphabets). In the case of POTOP, both pronunciations are legal Serbo-Croatian words (the former means "flood" and the latter, "rotor").

When a phonologically bivalent word of Serbo-Croatian is read in isolation, the alphabet is not specified by a context and, therefore, the spelled form can be pronounced in two different ways. Two types of such bivalent strings exist. In one of these, both pronunciations are known to the reader as words (i.e., have lexical entries). Such letter strings are both phonologically *and* lexically ambiguous. In the other type, which occurs more frequently, only one of the two pronunciations is a word, while the other is a nonword. Such strings are phonologically ambiguous, but since they are related to only one lexical entry they are not lexically ambiguous. An illustration of the various forms of ambiguity in Serbo-Croatian is presented on the lower portion of Figure 1.

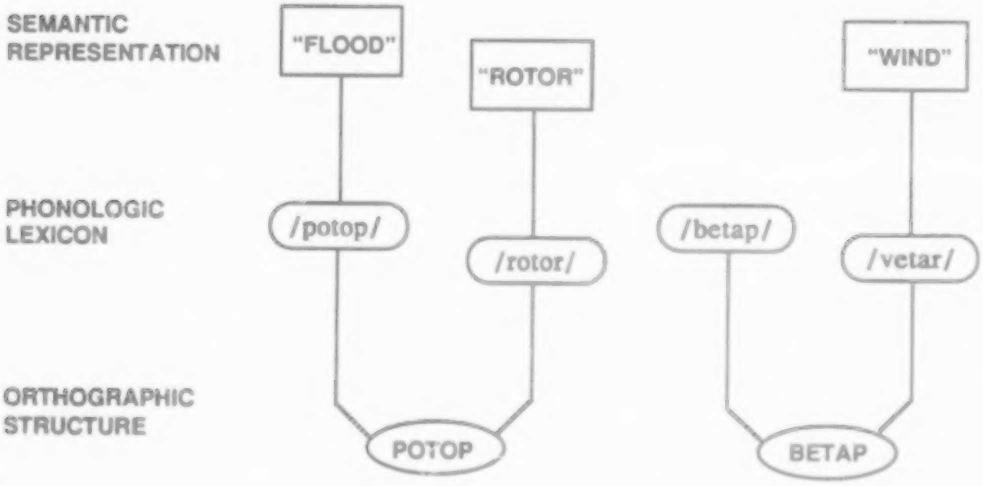
In the present study we exploited the bi-alphabetical property of Serbo-Croatian, in order to assess the relative contributions of phonological and lexical ambiguity to interference in the process of word recognition. Specifically, our aim was to investigate whether there are conditions under which phonological ambiguity in the absence of lexical ambiguity is sufficient to impair performance, or whether lexical ambiguity is a necessary condition for such interference, when words are presented in isolation. This issue may have relevance, at least to some extent, to orthographies which represent phonology in an indirect manner (i.e., printed letter strings can be read in more than one way). It focuses on whether phonologically ambiguous letter strings are processed differently regardless of the lexical status of the possible pronunciations.

Previous studies in Serbo-Croatian examined how fluent bi-alphabetic readers process ambiguous print. Lukatela, Popadić, Ognjenović, and Turvey (1980), investigated lexical decision performance in Serbo-Croatian, comparing phonologically ambiguous and unequivocal words. They demonstrated that words that could be pronounced in two different ways were accepted more slowly as words, relative to words that could be read in only one way. Similar results were found by Feldman and Turvey (1983) who compared phonologically ambiguous and phonologically unequivocal forms of the same lexical items. This outcome was interpreted to suggest that, in contrast to English, lexical decisions in Serbo-Croatian are necessarily based on the extraction of phonology from print (Turvey, Feldman, & Lukatela, 1984).

The relative contributions of phonological and lexical factors were not directly assessed in these previous studies. However, although Lukatela et al. (1980) demonstrated that phonologically ambiguous letter strings incurred longer lexical decision latencies than phonologically unequivocal strings, they did not find a significant difference in decision latencies between ambiguous strings with one or with two lexical entries. In fact, phonologically bivalent letter strings slowed subjects responses (although to a lesser extent), even if the two possible readings of the letter strings represented two nonwords (see also Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978). Interestingly, words that were composed exclusively of COMMON letters which were alphabetically bivalent but *phonologically unequivocal,* did not slow lexical decisions, relative to their unique alphabet controls (Feldman & Turvey, 1983). Finally, the magnitude of the difference in decision latencies to bivalent and unequivocal forms of a word varied with the number of ambiguous letters in the bivalent form of that word (Feldman, Kostić, Lukatela, & Turvey, 1983; Feldman & Turvey, 1983). In general, these results suggested that the reader of Serbo-Croatian processes print by a phonologically analytic strategy, that precedes lexical access. Consequently, as a rule, his or her performance is hindered by phonological ambiguity (See Feldman (1987), for a review).



ENGLISH



SERBO-CROATIAN

Figure 1. Different types of ambiguity in English and in Serbo-Croatian.

However, the conclusion that the phonologically analytic Serbo-Croatian reader is not sensitive to the lexical status of the different alphabetic readings may be premature. All of the above studies monitored either lexical decisions or naming of isolated words. These tasks may not have been sensitive enough to detect lexical effects. In particular, they may not have detected differences in the processing of phonologically ambiguous letter strings which represented one as contrasted with two words. Lexical decisions of isolated words can be made on a fast familiarity judgment (whether visual or phonological), that is based on a pre-lexical superficial recognition process, not necessarily involving lexical access (see Balota & Chumbley, 1984, 1985, for a discussion of this point). In contrast to lexical decision, the naming task may require a deeper phonological processing, but it does not necessarily reveal whether an alternative word, other than the reader's final choice, had been considered during the process of phonological disambiguation.

In a recent study Lukatela, Turvey, Feldman, Carello, and Katz (1989), examined whether a presentation of semantic context can override an ambiguous alphabetic context. They have shown that the correct and consistent alphabetic assignment for a letter string can indeed be offset by previously accessed lexical entries, (that are activated by semantic information). Hence, in contrast to the previous conclusions of Turvey et al. (1984), who suggested that a pre-lexical phonological analysis of print is mandatory in Serbo-Croatian, these results demonstrate that the Serbo-Croatian reader can also be affected by the lexical characteristics of the printed stimulus, if the experimental conditions invite deeper processing. For the purpose of the present study, in order to assess the relative effects of lexical and phonological ambiguity on the process of word recognition, we needed a methodology that tapped more directly the process of mapping printed words into lexical phonological structures. Therefore, in the present experiments we employed the "matching task" which was used in some of our previous studies for a similar purpose (Frost & Katz, 1989, Frost, Repp, & Katz, 1988).

The matching task

In the matching task subjects are simultaneously presented with a printed word (or a nonword) on a computer screen, and with a spoken word (or a nonword) via headphones. The subject is asked to decide as fast as possible whether or not the stimuli presented in the visual and the auditory modalities are the same (positive response), or whether they are different (negative response). Our assumption is that the simultaneous presentation of visual and auditory words produces separate activations of the orthographic and the phonologic lexical systems, that are, at least, *initially* independent. In order to match the spoken and the printed forms of words, they both have to be represented on the same metric, that is to converge at an identical lexical entry. Theoretically, this entry can be either phonologic or orthographic. Thus, the matching of a spoken and a printed word can be achieved, either by recoding the print into a phonological representation, and by comparing it to the phonological representation derived by the auditory channel, or, alternatively, by recoding the spoken form into an orthographic code and by comparing it to the printed word. However, these two theoretical possibilities are not equally probable. The transformation of speech into an orthographic representation is, by far, less practiced than the transformation of spelling into phonology. Indeed, in a previous study which involved simultaneous presentation of print and speech, we have demonstrated that it is the printed word which is transformed into a detailed phonetic representation, and that this process is very fast and automatic (Frost et al., 1988). Consequently, we assume that the common end result of both print and speech processing in the matching task is a *phonological* representation in the lexicon. The claim that the phonological representation is indeed lexical, is supported by results in a previous study that used the matching task (in English and Serbo-Croatian), in which word frequency effects were demonstrated, and differences were found between words and pseudowords (Frost & Katz, 1989).

The matching of an orthographic form to a spoken word might be performed in at least two different ways. One is to construct, at a first stage, functionally independent and complete phonological representations of the printed word and the spoken stimulus. These two representations are further compared at a subsequent stage. This possibility is in accordance with Morton's separate auditory and visual input logogens (Morton, 1981). A second possibility is that the process of mapping the print into phonology and the process of mapping the spoken stimulus into the common phonological structure are weakly interactive. According to this alternative, the two processes share information before completing their respective analyses, and by doing so, alter the extent of each other's processing. Thus, seeing print and forming a partial phonological representation of it can confirm (or contradict) a match to the accumulating phonological information being received via speech. In a complementary way, the rapid accumulation of information about which phonemes could be represented by the print can be influenced by the phonological information that is being received simultaneously via the auditory channel. Consistent with the interactive view are results that show reduced but significant cross-modal priming of auditory recognition of printed words (Hanson, 1980; Kirsner, Milech, & Standen; 1983), and visual-auditory interference in a Stroop paradigm in the naming task (Tanenhaus, Flanigan, & Seidenberg, 1980). It is important to distinguish between these two alternative possibilities in order to understand how phonology is derived from print.

In the three experiments reported here, subjects were presented with printed and spoken words, and were required to determine whether they are equivalent. The printed words were selected for the matching task so that in their printed form they were phonologically ambiguous in one condition (e.g., "POTOP"), and unequivocal in another condition (an unambiguous form of the same word, e.g., "ROTOR," "ПОТОП"). The spoken forms of the words presented auditorily were always (necessarily) unambiguous, (e.g., /rotor/). Our primary aim was to determine the effects of phonological and lexical ambiguity. If phonological ambiguity in itself, affects the process of word recognition, then the efficiency of the matching process should be directly influenced by the number of possible pronunciations for the printed word, regardless of the lexical status of the different pronunciations. If lexical ambiguity (whether a letter string has two meaningful pronunciations or only one) affects the matching process, then subjects' performance for phonologically ambiguous strings that are meaningful in both their Cyrillic and their Roman forms, should be slower than phonologically ambiguous strings that are meaningful in only one alphabet. In both instances, performance will be assessed by computing the difference between latencies for matching the spoken words to the bivalent and unequivocal printed forms of the same word. An additional aim of the present study was to examine the sequence of the early stages of the matching process. Hence, we introduced an additional experimental manipulation which consisted of slowing the process of generating phonological information from speech by degrading the spoken presentation.

EXPERIMENT 1

In Experiment 1 we presented subjects with phonologically bivalent letter strings that could be read as a word or as a nonword, and with bivalent strings that could be read as two legal words (the former strings were phonologically but not lexically ambiguous, while the latter were ambiguous at both the phonological *and* the lexical level). In the present experiment, when a lexically ambiguous letter string was presented in print, its *high-frequency* (dominant) reading was always presented auditorily as the matching referent. Similarly, the dominant alternative was always depicted when the unequivocal printed form of the same letter string appeared. The purpose of the experiment was to investigate whether lexical ambiguity is indeed perceived and affects the reader's performance even when the dominant reading of the bivalent letter string is presented.

A second aim of the present study was to investigate the time course of processing the printed and the spoken word, by examining the interaction of auditory degradation with phonological and lexical ambiguity. If the subject derives the phonological representation of the auditory input independently of information coming from the visual channel, then the effect of auditory degradation should be equivalent when the printed stimulus is phonologically ambiguous and when it is not; in other words, the effects of degradation and ambiguity should be additive. Auditory degradation will slow the process of retrieving the phonological structure from the phonetic input, but this reduction in the amount of phonological information will have no effect on the generation of phonology from the print. On the other hand, if the partial phonemic information arriving on the auditory channel facilitates the selection of a phonemic value for those graphemes that are phonologically ambiguous (before the spoken word is fully analyzed), then the effects of degradation and phonological ambiguity should interact: The matching of letters to their respective phonemes will be slower for the phonologically ambiguous letters than for the unequivocal letters (see Frost and Katz (1989); Frost, Repp, and Katz (1988) for a discussion of the simultaneous presentation with degradation technique).

Method

Subjects

Forty undergraduate students from the University of Belgrade, all native speakers of Serbo-Croatian, participated in the experiment for course credit.

Stimuli and Design

A total of 120 pairs of words were used in the experiment. A pair consisted of a visual and an auditory presentation of a word. In sixty pairs, the printed and the spoken words were the same, and in sixty pairs they were different. The 60 "same" pairs included our target words, while the 60 "different" pairs were introduced as fillers in order to achieve a probability of 0.5 for a "same" response.

The printed words were presented in Cyrillic or Roman characters, and appeared either in a phonologically bivalent form or in a phonologically unequivocal form. Of the phonologically bivalent target words, 20 had two meaningful pronunciations, one in each alphabet (usually belonging to the same syntactic class), and 20 had only one meaningful pronunciation, half Roman and half Cyrillic, the other being a nonword. In addition there were 10 unequivocal fillers in each alphabet. In summary, the Roman and the Cyrillic scripts were mixed equally in all experimental conditions.

Obviously, the auditory form was always phonologically unequivocal. For the lexically ambiguous letter strings (i.e., those that had two possible meaningful pronunciations), the spoken form was always the *higher-frequency* (dominant) lexical alternative. Since there is no standard frequency count in Serbo-Croatian, the relative frequencies of the two pronunciations were assessed by 30 native speakers. They were presented with the phonologically unequivocal printed forms of the high- and the low-frequency alternatives, and were requested to decide for each pair which alternative was more frequent. Based on this judging procedure, we selected 20 ambiguous words for which there was at least 75% agreement across judges on which alternative was the dominant meaning. Thus, the auditory stimulus always matched the higher frequency alternative of the printed bivalent word. For the phonologically bivalent but lexically unequivocal letter strings, the spoken form always was the lexical (meaningful) alternative, and not the nonword. The experimental design is summarized in Table 1.

Mismatched trials were constructed by pairing two words with the same length, and vowel-consonant structure, but which differed with respect to one or two phonemes. The position of the nonidentical segments was distributed across words uniformly. Half of the visually presented trials were in Roman and half in Cyrillic. Auditory degradation was introduced in the experiment as an additional factor. Each subject heard an equal number of degraded and undegraded stimuli, and this condition was counterbalanced across subjects.

Four test lists were created each containing all four combinations of ambiguity and degradation: For each subject, a word could appear in either its phonologically ambiguous form or its unequivocal form with an auditory version that either was or was not degraded by noise. Each test list contained all four combinations of different items, and across lists a given word appeared in each of the four conditions. Each subject was tested with only one list. This design allowed us first, to compare the effects of phonological ambiguity (one or two possible pronunciations), to the effects of lexical ambiguity (one or two meaningful pronunciations of the printed word), and second, to examine whether auditory degradation interacted with either phonological ambiguity or lexical ambiguity.

The visual stimuli were presented on a Macintosh computer screen. They subtended a visual angle of approximately 2.5 degrees on average. The auditory stimuli were originally spoken by a male native speaker in an acoustically shielded booth and recorded on an Otari MX5050 tape-recorder. The speech was digitized at a 20 kHz sampling rate. Each digitized stimulus was edited using a digital waveform editor.

Table 1. *Examples of words in each experimental condition. Note, that the Roman reading of BETAP is /betap/ which is a nonword. In contrast the Cyrillic reading of POTOP is /rotor/ which is indeed a word.*

ONE MEANINGFUL PRONUNCIATION			TWO MEANINGFUL PRONUNCIATIONS		
Unequivocal Print	Bivalent Print	Auditory Stimulus	Unequivocal print	Bivalent Print	Auditory Stimulus
VETAR	BETAP	/vetar/	потоп	POTOP	/potop/

Its onset was determined visually on an oscilloscope, and was verified auditorily through headphones. A mark tone was then inserted, at the onset of each stimulus. The digitized edited stimuli were recorded at three-second intervals on a two track audiotape, one track contained the spoken words while the other track contained the mark tones. The purpose of the mark tone was to trigger the presentation of the printed word on the computer screen.

Auditory degradation was achieved by masking each spoken stimulus with signal-correlated noise—that is, white noise with the same amplitude envelope as the spoken stimulus (Schroeder, 1968). Speech and noise were added digitally in proportions of 0.3 and 0.7, corresponding to a constant signal-to-noise ratio of -7.5 dB. The use of signal-correlated noise insured that the characteristics of the auditory degradation were equivalent for all of the stimuli.

Procedure and apparatus

Subjects wore headphones and sat in a semi-darkened room in front of the computer screen. The experimental task consisted of pressing a "Same" key if the visual and the auditory stimuli were the same word, and pressing a "Different" key if they were different. The dominant hand was always used for the"yes" response.

The audiotape containing the verbal stimuli and the mark tones was played on a two channel Sony tape recorder. The verbal stimuli were transmitted binaurally to the subject's headphones. The mark tones were transmitted through an interface to the Macintosh, where they triggered the visual presentation and the computer's clock for reaction time (RT) measurements. The experimental session began with 16 practice pairs. After the practice, all 120 test trials were presented in one block.

Results and Discussion

Means and standard deviations of RTs for correct responses were calculated for each subject in each of the experimental conditions. Within each subject-condition combination, outliers that were greater than two standard deviations from the mean were eliminated. Outliers accounted for less then five percent of all responses. This procedure was repeated in all three experiments of the present study.

For the sake of clarity we will hereby define the various effects which were assessed in our analyses. 1) The main effect of phonological ambiguity was measured by comparing RTs to words which were presented in their bivalent alphabet, with RTs to the same words when presented in their unequivocal alphabet. 2) The main effect of lexical ambiguity reflects differing decision latencies to words which have two meaningful pronunciations in their bivalent form, and words which have only one meaningful pronunciation. Note that this main effect is difficult to interpret, since we are dealing here with two groups of words which differ in their respective frequencies. We will, therefore, ignore this main effect in our following analyses. A more meaningful comparison is 3) the interaction of lexical ambiguity, defined in terms of number of meanings, and phonological ambiguity, assessed in terms of the difference between bivalent and unequivocal forms. This interaction would suggest that phonological ambiguity is indeed more detrimental for words which have two meaningful pronunciations than for words which have only one. 4) The main effect of degradation reflects the different RTs for degraded and undegraded auditory stimuli. This main effect is, in fact, trivial. Of more interest is 5) the interaction of degradation with phonological ambiguity and lexical ambiguity.

RTs in the different experimental conditions are summarized in Table 2. Overall, RT was slower when words were printed in their phonologically ambiguous form than in their unequivocal form. Moreover, the difference between ambiguous and unequivocal forms appeared to be larger when the bivalent words had two meaningful pronunciations, rather than one. Finally, the effects of auditory degradation were quite similar in all experimental conditions.

To assess the statistical significance of these trends we performed an analysis of variance across subjects (F1) and across stimuli (F2), with the main factors of phonological ambiguity (unequivocal or phonologically bivalent print), lexical ambiguity (one meaningful pronunciation or two), and degradation (normal or degraded presentation). The main effects of phonological ambiguity and degradation were significant: F1 (1,39) = 61.3, MSe = 164,054, $p < 0.001$; F2 (1,19) = 26.2, MSe = 96,485, $p < 0.001$ and F1 (1,39) = 20.7, MSe = 234,155, $p < 0.001$; F2 (1,19) = 18.2, MSe = 87,701, $p < 0.001$, respectively. The interaction of lexical ambiguity and phonological ambiguity was marginally significant: F1(1,39) = 3.07, MSe = 207,802, $p < 0.08$; F2 (1,19) = 2.05, MSe = 138,633, $p < 0.1$. The effect of degradation did not interact with either print or lexical ambiguity (F1, F2 < 1.0). The three way interaction was also nonsignificant (F1, F2 < 1.0).

Table 2. *Matching RTs (in ms) and Errors (in parenthesis) for words printed in unequivocal or bivalent print, under normal or degraded auditory presentations. Bivalent words had either one or two lexical entries in their bivalent forms. The dominant alternative was presented auditorily.*

	ONE MEANINGFUL PRONUNCIATION			TWO MEANINGFUL PRONUNCIATIONS		
	Unequivocal Print	Bivalent Print	Bivalence Effect	Unequivocal print	Bivalent Print	Bivalence Effect
Normal	740	787	47	707	778	71
	(4%)	(5%)		(5%)	(12%)	
Degraded	780	817	37	751	823	72
	(8%)	(7%)		(11%)	(17%)	

We analyzed the percentage of errors in a separate analysis of variance. The results are similar to the RT analysis and are presented in Table 2. The main effects of phonological ambiguity and degradation were significant: F1(1,39) = 3.6, MSe = 24.6, $p < 0.06$; F2 (1,19) = 5.10, MSe = 17.5, $p < 0.03$; and F1(1,39) = 11.0, MSe = 12.0, $p < 0.001$; F2(1,19) = 8.65, MSe = 10.5, $p < 0.008$; respectively. The interaction of phonological ambiguity and lexical ambiguity was marginally significant in the subject analysis, F1(1,39) = 3.40, MSe = 26.1, $p < 0.07$, but highly significant in the stimuli analysis, F2 (1,19) = 8.65, MSe = 13.6, $p < 0.008$. Again, degradation did not interact with the other factors.

Although the results were not fully consistent, they suggest that phonological ambiguity affected word recognition: If a letter string could be pronounced in two ways, subjects' performance was slowed, even when only one of those pronunciations was a real word. It is important to emphasize that because the unequivocal and ambiguous forms were related to the same word, differences between these two forms cannot be related to differences in their relative frequency. If anything, the cumulative frequency for the bivalent form was higher than for the unequivocal form (see also Feldman et al., 1983). The effect of phonological ambiguity appeared to be even larger when the ambiguous printed stimulus could have been read as two meaningful words than when it could be read in only one meaningful way. However, this interaction was more pronounced in the error analysis.

The effects of auditory degradation provide additional insights concerning the interaction of the visual and the auditory systems in word recognition. Clearly, auditory degradation of the spoken stimulus resulted in an overall slowing of subjects' responses. However, this effect was additive, and did not interact with phonological or lexical ambiguity. This result suggests that, at least in the processing of high-frequency words, the extraction of phonemic information from the auditory signal is performed independently of the visually-based processing of the words presented in print. In Experiment 2, we determine if this outcome can be generalized to the perception of the low-frequency alternative reading of bivalent words.

EXPERIMENT 2

The effects of lexical ambiguity revealed in Experiment 1 may have been attenuated because only the high-frequency lexical alternative of a phonologically bivalent word was presented to the subject. It is possible that when the lexical alternatives of a bivalent letter string are discrepant in frequency, only the dominant alternative is activated sufficiently, and therefore lexical ambiguity does not have a strong effect on the process of word recognition. This could have resulted in a considerable reduction of the ambiguity effect, thereby attenuating the interaction between phonological ambiguity and lexical ambiguity observed in Experiment 1.

Our aim in Experiment 2 was to examine the effect of phonological ambiguity and its interaction with lexical ambiguity and auditory degradation when the *low-frequency* (subordinate) alternatives are presented as unequivocal controls.

Method

Subjects

Forty undergraduate students from the University of Belgrade, all native speakers of Serbo-Croatian, participated in the experiment for credit. None of them had participated in Experiment 1.

Stimuli and Design

The stimuli, design, procedure, and apparatus were identical to those used in Experiment 1, except that when the bivalent words had two meaningful pronunciations, the *low-frequency* meaning was presented auditorily as the matching referent, and as the unequivocal printed control.

Results and Discussion

RTs and errors in the different experimental conditions are summarized in Table 3. As in Experiment 1, we performed an analysis of variance across subjects (F1) and across stimuli (F2), to assess the relative effects of phonological ambiguity, lexical ambiguity, and auditory degradation.

The main effects of phonological ambiguity and auditory degradation were significant: F1(1,39) = 115, MSe = 295,386, $p < 0.001$, F2 (1,19) = 96.3, MSe = 74,130, $p < 0.001$; F1(1,39) = 37.7, MSe = 258,479, $p < 0.001$, F2 (1,19) = 8.44, MSe = 286,057, $p < 0.01$, respectively. The interaction of phonological ambiguity and lexical ambiguity was highly significant: F1(1,39) = 55.2, MSe = 145,027, $p < 0.001$, F2 (1,19) = 18.1, MSe = 94,976, $p < 0.001$. Finally, as in Experiment 1, degradation did not interact either with phonological ambiguity (F1, F2 < 1), or with lexical ambiguity (F1< 1.0, F2 = 1.6).

Similar results were obtained in the error analysis. There was a significant main effect of phonological ambiguity, F1(1,39) = 28.2, MSe = 9.1, $p < 0.001$, F2(1,19) = 13.0, MSe = 7.0. $p < 0.001$, and auditory degradation, F1(1,39) = 23.5, MSe = 7.8, $p < 0.001$, F2 (1,19) = 20.7, MSe = 10.0, $p < 0.001$. There was a significant interaction of phonological ambiguity and lexical ambiguity F1(1,39)= 4.87, MSe = 10.0, $p < 0.03$. Degradation did not interact with phonological ambiguity (F1 < 1.0, F2 =1.97), or with lexical ambiguity F1 (1,39) = 3.5, MSe = 16.8, $p < 0.07$, F2 (1,19) < 1.0, MSe = 9.7, $p < 0.4$).

Thus, the trends revealed in Experiment 1 were replicated in Experiment 2, but the differences were strongly enhanced. The effect of phonological ambiguity was three times larger for words that had two meaningful pronunciations, than for words with only one. When this interaction is compared to the one obtained in Experiment 1, it is clear that the major difference occurs for stimuli that could be read as two legal words. Here, the effect of phonological ambiguity was twice as large as in Experiment 1 (approximately 140 ms, as compared to 70 ms in the previous experiment). The statistical significance of this difference was assessed by using a mixed model ANOVA design, with the additional between-subjects factor of frequency. The three-way interaction of phonological ambiguity, lexical ambiguity, and frequency, was indeed significant F1(1,78) = 11.7, MSe = 4523, $p < 0.001$.

In spite of the enhanced effects of phonological ambiguity and its interaction with lexical ambiguity, it is clear that the effect of auditory degradation was similar for phonologically bivalent strings with one and two meaningful pronunciations, and for phonologically unequivocal words. This pattern was revealed in both the RT and error analyses and was, in fact, very similar to the pattern obtained in Experiment 1. This outcome is consistent with our claim that the processing of the degraded auditory information was completed independently of the visual input.

Table 3. *RTs (in ms) and Errors (in parenthesis) when the subordinate alternative was presented auditorily.*

	ONE MEANINGFUL PRONUNCIATION			TWO MEANINGFUL PRONUNCIATIONS		
	Unequivocal Print	Bivalent Print	Bivalence Effect	Unequivocal print	Bivalent Print	Bivalence Effect
Normal	749	801	52	749	894	145
	(1%)	(3%)		(10%)	(14%)	
Degraded	808	850	42	818	970	152
	(5%)	(10%)		(13%)	(25%)	

Our results seem to suggest that a phonologically bivalent letter string was first recoded as (or activated) its dominant phonological alternative which was then compared to the spoken word. Since, in the present experiment, the spoken word represented the subordinate alternative, the subject would first find a mismatch and only then would generate the less frequent alternative. This tendency might explain the poorer overall performance in the present experiment.

EXPERIMENT 3

In both Experiments 1 and 2, we found that the degradation of the auditory input did not interact with phonological ambiguity or with lexical ambiguity. These results lead us to conclude that subjects constructed complete phonological representations of the printed and the spoken stimuli, independently of each other. It is possible, however, that this outcome resulted from different time courses of lexical activation caused by the printed and the spoken words. Although the onsets of the visual and auditory presentations were simultaneous, all of the print information was available immediately, while the auditory information that constituted the spoken word was distributed over 400 to 500 ms of time. Thus, in Experiments 1 and 2, the auditory information effectively lagged behind the visual information. This basic difference between auditory and visual presentation could have caused the subject to process the print first and the speech subsequently and independently. A similar argument was proposed by Glucksberg, Kreuz, and Rho (1986), who suggested that in a cross-modal backward priming paradigm, a visual target word presented after the onset of an auditory prime word might affect the lexical interpretation of the auditory stimulus. The purpose of Experiment 3 was to investigate whether auditory degradation did not interact with phonological ambiguity because of the particular timing relation between visual and auditory stimulus presentations, or whether the outcome we observed indeed reflects a more general strategy for processing bi-modal verbal information.

In Experiment 3, the visual presentation was delayed by 200 ms, relative to the onset of the spoken word. (A lag of 200 ms is slightly shorter than the average recognition point of a two-syllable word presented auditorily (e.g., Marslen Wilson & Tyler, 1980). Thus, the printed word was presented not at the onset of the spoken word, but *during* its distributed presentation. If the subject had the tendency, in the first two experiments, to process the print first, the new timing relation should counteract this bias. As noted above, the purpose of this manipulation was to encourage parallel processing of print and speech, in order to increase the possibility of interactive processing.

Method

Subjects

Eighty undergraduate students from the University of Belgrade, all native speakers of Serbo-Croatian, participated in the experiment for credit. None of the subjects participated in the previous experiments. Forty of the 80 subjects (group A) were tested with the stimuli of Experiment 1 (dominant alternatives), and the other 40 (group B) were tested with the stimuli employed in Experiment 2 (subordinate alternatives).

Stimuli, Design, and Procedure

The stimuli, design, and apparatus were identical to those used in Experiments 1 and 2. The procedure, however, differed in one important respect from the previous experiments: the visual presentation of the stimuli lagged from the speech onset by 200 ms, and the frequency of the lexical alternatives for spoken forms was manipulated between sets of subjects.

Results and Discussion

RTs are summarized in Table 4. We conducted separate analyses of variance for subjects who were presented with the high-frequency spoken alternatives (Experiment 3a), and for subjects who were presented with the low-frequency spoken alternatives (Experiment 3b). Each analysis was performed across subjects (F1) and across stimuli (F2), with the main factors of phonological ambiguity, lexical ambiguity, and auditory degradation.

For the group presented with the dominant alternatives (group A), the main effects of phonological ambiguity and degradation were significant: F1 (1,39) = 55.4, MSe = 176,745, $p < 0.001$, F2 (1,19) = 20.7, MSe = 128,627, $p < 0.001$, and F1 (1,39) = 25.6, MSe = 382,340, $p < 0.001$, F2 (1,19) = 22.7, MSe = 121,110, $p < 0.001$. The interaction of phonological ambiguity and lexical ambiguity was significant, F1 (1,39) = 13.1, MSe = 170,770, $p < 0.001$, F2 (1,19) = 4.37, MSe = 153,201, $p < 0.05$. More importantly, just as in Experiments 1 and 2, degradation did not interact with lexical ambiguity, or with phonological ambiguity (F1, F2 < 1.0).

Table 4. *RTs (in ms) and Errors (in parenthesis) when the onset of the dominant (3a) and the subordinate (3b) alternatives was presented auditorily before the visual form.*

DOMINANT

	ONE MEANINGFUL PRONUNCIATION			TWO MEANINGFUL PRONUNCIATIONS		
	Unequivocal Print	Bivalent Print	Bivalence Effect	Unequivocal print	Bivalent Print	Bivalence Effect
Normal	637	666	29	580	658	78
	(1%)	(1%)		(6%)	(9%)	
Degraded	688	718	30	636	724	88
	(13%)	(11%)		(14%)	(22%)	

SUBORDINATE

	ONE MEANINGFUL PRONUNCIATION			TWO MEANINGFUL PRONUNCIATIONS		
	Unequivocal Print	Bivalent Print	Bivalence Effect	Unequivocal print	Bivalent Print	Bivalence Effect
Normal	615	652	37	621	713	92
	(1%)	(3%)		(10%)	(17%)	
Degraded	688	723	35	700	781	81
	(11%)	(17%)		(20%)	(30%)	

The error analysis revealed a significant effect of auditory degradation, F1(1,39) = 40.0, MSe = 22.7, $p < 0.001$, F2 (1,19) = 46.7, MSe = 20.0, $p < 0.001$, and a significant interaction of phonological ambiguity and lexical ambiguity, F1(1,39) = 5.9, MSe = 12.9, $p < 0.01$, F2 (1,19) = 3.72, MSe = 20.8, $p < 0.06$. No interaction with degradation was found (F1, F2 < 1.0).

Similar results were obtained with the experimental group presented with the subordinate alternatives (group B): all main effects were significant: F1(1,39) = 41.6, MSe = 281,833, $p < 0.001$, F2 (1,19) = 17.3, MSe = 165,043, $p < 0.001$; F1(1,39) = 36.4, MSe = 455,213, $p < 0.001$, F2 (1,19) = 38.7, MSe = 118,930, $p < 0.001$; for phonological ambiguity and auditory degradation, respectively. The interaction of phonological ambiguity and lexical ambiguity was significant: F1(1,39) = 7.93, MSe = 251,528, $p < 0.007$, F2 (1,19) = 5.13, MSe = 84,984, $p < 0.03$. As in all previous experiments, degradation did not interact with either phonological ambiguity, (F1, F2 < 1.0), or with lexical ambiguity, (F1, F2 = 0.0), nor was the third-order interaction significant (F1, F2 < 1.0).

The error analysis revealed only significant main effects for phonological ambiguity, F1(1,39) = 28.9, MSe = 11.8, $p < 0.001$, F2 (1,19) = 5.53, MSe = 26.0, $p < 0.02$, lexical ambiguity, F1(1,39) =69.0, MSe =15.9, $p < 0.001$, F2 (1,19) = 5.1, MSe =35.7, $p< 0.03$, and auditory degradation, F1(1,39) = 45.6, MSe = 23.1, $p < 0.001$, F2 (1,19) = 35.2, MSe = 26.1, $p < 0.001$.

The results obtained in Experiment 3 are very similar to those obtained in Experiment 1 and 2. In general, delaying the printed information did not interfere with subjects' performance. Similar error rates were found for normal and delayed presentation, and the faster RTs in the delayed condition are, in fact, an artifact of delayed measurement onsets (In all experiments, RTs were measured from print onset). Regardless of which phonological alternative was presented auditorily, phonological ambiguity did affect subjects' performance. Moreover, the effect, was stronger when the phonologically ambiguous words had two meaningful pronunciations than when they had only one. If we consider first, the results obtained with the dominant alternatives, it appears that our experimental manipulation actually enhanced the interaction between phonological ambiguity and lexical ambiguity, obtained in Experiment 1. Whereas the statistical significance of these trends was unclear in Experiment 1, Experiment 3a provides a reliable confirmation of this observed effect. Even when the subject hears the dominant alternative of a phonologically bivalent word, *both* lexical (and phonological) interpretations are automatically activated. Nevertheless, in spite of the enhanced effect found in Experiment 3a, degradation still did not interact with either phonological ambiguity or with lexical ambiguity. We conclude, again, that the two processes of generating phonological representations from the visual and auditory input channels were independent.

The results obtained with the delayed visual presentation of the subordinate alternatives (Experiment 3b), suggest that the effect of phonological ambiguity for words with two meaningful pronunciations was reduced relative to the effect we found in Experiment 2. In order to assess the statistical significance of this difference, we conducted an analysis which compared directly the interaction of phonological and lexical ambiguity in Experiments 2 and 3b. We used a mixed ANOVA design with the additional between subjects factor of lag. The analysis revealed that the three-way interaction of phonological ambiguity, lexical ambiguity, and lag, was indeed significant ($F(1,78) = 5.1$, $MSe = 5084$, $p < 0.02$). This outcome provides additional insights concerning the effect of lagged visual presentation on the disambiguation process. We will discuss this finding further in the General Discussion.

GENERAL DISCUSSION

The present study was designed to examine the effects of phonological ambiguity and lexical ambiguity in print on word recognition. In three experiments, we exploited a special property of the Serbo-Croatian language—its transcription to both the Roman and Cyrillic alphabet—in order to manipulate separately the phonological ambiguity of a printed word and the number of lexical entries it represents. Our subjects were presented simultaneously with a printed and a spoken word and were required to determine whether they were the same or different. The speed and the accuracy of matching the printed and the spoken forms were measured when the print had either one possible pronunciation or two, and if two, when both pronunciations were real words or only one was. In addition we examined whether auditorily degrading the spoken word interacted with either lexical or phonological ambiguity.

Across experiments, we manipulated the frequency of the lexical alternatives represented by the phonologically bivalent word, as well as the temporal interval between the onset of the speech and the print. A consistent pattern of results emerged throughout the study. First, both the number of possible pronunciations for a printed word and their lexical status affected word recognition. Second, these effects interacted: phonological ambiguity had a greater effect when the bivalent form could be read as two legal words than when it could be read as only one. Third, auditory degradation never interacted either with print ambiguity or with lexical ambiguity. In addition, these effects were obtained regardless of the lexical alternative's frequency or the temporal onset of print presentation.

Previous studies using the lexical decision task examined how fluent bi-alphabetic readers of Serbo-Croatian process ambiguous print. These studies showed that phonologically bivalent letter strings, both words and nonwords, indeed incurred longer decision latencies than different unequivocal strings. However, the effect of phonological ambiguity was similar for phonologically bivalent letter strings whether the string represented real words or nonwords (Lukatela et al. 1978; 1980).[1] In contrast to these previous studies, our results suggest that the reader is affected not only by the number of possible pronunciations of a printed word, but also by their lexical status. That is, print bivalence caused greater difficulties when both of the pronunciations were attached to lexical entries than when only one pronunciation was a real word. One advantage of the design employed in the present study is that it allowed us to demonstrate the relative contribution of lexical and phonological factors, by comparing subjects' performance for the *same* word presented in either a bivalent printed form or in an unequivocal printed form. We found, in all three experiments, that the inhibitory effect on reaction time caused by bivalent print was roughly twice as large for bivalent strings with two lexical entries as compared to bivalent strings with only one real word entry (the other pronunciation being a nonword).

As noted above, the discrepancy in results between the present and previous studies may be due to the different tasks that were employed. The matching task explicitly requires matching the printed and the spoken forms of a word. Effects of lexical ambiguity should arise whenever the printed letter string is recoded into more than one phonological form. In this case two conflicting alternative representations will be matched with the phonological representation derived from the auditory input which is necessarily unambiguous. In contrast, lexical decisions can be made on a fast familiarity judgment (Balota & Chumbley, 1984, 1985), and naming can be performed by a selection of only one lexical candidate. One consequence of the different task requirements is that lexical decision and naming can be performed successfully on the basis of one complete representation. Therefore, our results seem to provide evidence that the matching task is more sensitive than the lexical decision and naming tasks for detecting effects of phonologically-derived lexical ambiguity.

A large effect of lexical status was obtained in the present study: phonological ambiguity hindered performance to a greater extent when both pronunciations were meaningful than when only one pronunciation was a word. This effect suggests that interference was caused by the conflict between the two lexical entries. Although lexical effects in phonological disambiguation have been demonstrated before in Serbo-Croatian, they were studied only in a priming paradigm. That is, a prime that was either semantically or alphabetically related to one of the lexical entries contributed to the disambiguation process (Feldman, 1987). Similarly, Lukatela et al. (1989) have shown that prior lexical access can override the consistent assignment of an alphabet reading in cases where the target letter string contained illegal combinations of Roman and Cyrillic letters. Our results with the matching task suggest that lexical factors can also be demonstrated when bivalent strings are presented in isolation.

The influence of frequency provides additional insights concerning the processing of bivalent letter strings. When the subordinate alternative was presented auditorily, it inhibited the response to its matching but bivalent printed letter string: In Experiment 2 and 3B (which used the low-frequency spoken words), the difference between bivalent and unequivocal printed forms was greater than was the difference in Experiments 1 and 3a. This outcome suggests that subjects initially access the dominant alternative related to the bivalent printed string, and that a mismatch between its phonological form and the subordinate form presented on the auditory channel was the cause for the longer RTs (and greater number of errors) encountered in these experimental conditions. These results are parallel to meaning frequency effects of homophonic homographs reported by Hogaboam and Perfetti (1975) and, overall, support frequency-ordered access models for lexically ambiguous words (e.g., Simpson, 1981; Simpson & Burgess, 1985).

In a task in which subjects had to find a polysemous word in a sentence, Hogaboam and Perfetti (1975) demonstrated that detection was faster when the context was related to the subordinate meaning of the ambiguous word than when it was related to the dominant meaning. Similar results have been recently reported by Neill, Hilliard, and Cooper (1988). In both studies, this outcome was interpreted to indicate that the more frequent or dominant meaning of a homograph was retrieved at an initial stage. Because the initially activated dominant meaning did not match the subordinate meaning which was suggested by the context, ambiguity was easily detected by the subjects. Evidence for the differential activation for high- and low- frequency meanings of ambiguous homographs was also demonstrated with event-related potentials (Van Petten & Kutas, 1987) and by monitoring eye movements (Duffy, Morris, & Rayner, 1988; Rayner & Frazier, 1989). These studies, however, examined the processing of homographs embedded in context. More relevant to our study are the results presented by Simpson and Burgess (1985) who examined the order of retrieving alternative meanings of *isolated* homographs. In their study, Simpson and Burgess found that the priority for accessing the alternative meanings, was determined primarily by their relative frequencies: The more frequent meaning was retrieved at an initial stage, and the subordinate meaning was activated subsequently. The above studies with English materials examined the time course of activation of homophonic homographs with two meanings and a single pronunciation. Perhaps, the same principle of frequency-ordered access can be extended to the materials in the present study: heterophonic homographs with two meanings and two pronunciations.

Note, however, that the effect of print bivalence was obtained even when only the dominant meaning was appropriate, and the subordinate meaning was irrelevant (Experiments 1 and 3a). Even though the spoken word in these experiments was the dominant meaning, we still found greater effects of print bivalence for words with two lexical entries, over words with one lexical entry. This result invalidates the possibility that only one meaning, the more frequent lexical alternative, was generated from the bivalent letter string. Rather, it suggests that the subordinate alternative was activated as well, even when the dominant alternative was presented auditorily as the matching referent. The activation of the subordinate alternative hindered subjects' performance for words with two lexical entries because two lexical entries had to be considered instead of one.

In general, this outcome suggests exhaustive access to all meanings of ambiguous words and that these entries are automatically activated in the process of word recognition (e.g., Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowsky, 1982). Our results further support a model in which both high- and low-frequency alternatives are indeed activated but not in parallel; their order of activation is determined by their relative frequency, (Duffy et al., 1988; Neill et. al., 1988).

It appears that the Serbo-Croatian reader processes print in a phonologically analytic manner. He/she is sensitive to the phonological ambiguity presented by the orthographic structure, and derives from it both phonological forms. Nevertheless, the dominant alternative usually has an earlier or higher level of activation and, therefore, predominates over the subordinate alternative. However, this dominance effect can be reduced or reversed by priming the subordinate alternative of the print with the spoken (unambiguous) equivalent. This interpretation is supported by the effects of delaying the onset of the print in Experiment 3b.

In the delayed presentation of the subordinate words, the ambiguity effect observed in Experiment 3b was reduced almost by half relative to the effect obtained in Experiment 2. This result demonstrates that the spoken word was at least partially analyzed before the bivalent word was presented. The partially perceived subordinate spoken word enhanced the level of activation of its lexical representation, effectively priming the subordinate lexical entry. Ordered access models assume different time courses of activation for the dominant and the subordinate alternatives of ambiguous words (e.g., Simpson & Burgess, 1985; Van Petten & Kutas, 1987). Simpson and Burgess (1985) have shown that the dominant meaning of a polysemous homograph was activated as soon as 16 ms from stimulus onset, whereas the subordinate meaning was activated only 100 ms from stimulus onset. Our results suggest that the partial presentation of the spoken word prior to the printed stimulus enhanced the activation of subordinate alternatives which was triggered by the print, thereby reducing the frequency effect found in Experiment 3b by half, relative to Experiment 2. However, the 200 ms by which the auditory information preceded the visual information did not completely resolve the printed word's phonological or lexical ambiguity: Ambiguity effects still remained in the delayed presentation. This outcome probably suggests that auditory lexical access was not completed within the 200 ms time period.

We now turn to the effects of auditory degradation. Superimposing noise on the auditory channel affected subjects' performance, but its effect was additive to the effects of phonological and lexical ambiguity, the spoken word's frequency, and the onset of print presentation. This outcome suggests that processing of the masked spoken words was independent of the printed information. Apparently, the built-up of phonological information necessary for the perception of the spoken word proceeded without reference to information derived from the printed form.

Recently, Frost and Katz (1989) proposed a model that describes the mental lexicon as containing two levels of interactive relationship between orthographic and phonological representations: One level contains the representations of letters and phonemes and relations between them. The other level involves whole printed and spoken words. The relations between spelling and sound in different orthographies (Serbo-Croatian as opposed to English, for example) was described in terms of the structural differences that exist among the various languages in respect to the relationship between the orthographic and phonological systems. In a similar way, phonological ambiguity can be characterized by complex, non-isomorphic connections between letters and phonemes, and lexical ambiguity can be characterized by ambiguous connections between the orthographic and the phonological systems at the level of whole words. The results of the present study provide additional information as to how the phonological and orthographic systems interact in a shallow orthography like Serbo-Croatian.

The additive effect of auditory degradation and print ambiguity suggests that the perception of spoken words was not penetrated by information from the orthographic system. When subjects heard degraded speech, the presence of potentially helpful (i.e. phonologically unambiguous) print did not, in fact, facilitate the match. Note, that since auditorily degraded and undegraded stimuli were randomly mixed, subjects could not have adopted one uniform strategy throughout the experimental session (e.g., favoring visual analysis when all stimuli were auditorily degraded). There are two possible explanations for the absence of such an interaction. Perhaps, our previous suggestion that the orthographic and phonological systems exchange information at the level of single graphemes and phonemes (Frost & Katz, 1989; Frost et al., 1988) should be constrained to deep orthographies—in which such an exchange might be beneficial. English could benefit from such interaction because in English some letters may represent more than one phoneme. Accordingly, the exchange of information between the visual and the auditory modalities at this level can provide the reader of English with additional information for resolving ambiguity. In contrast, in Serbo-Croatian, the simplicity of the relation between letter and phoneme may preclude any further improvement due to an interaction of printed and auditory information.

A second possibility for the absence of an interaction between phonological ambiguity and degradation is that the perception of spoken words may be isolated from information derived from orthographic processing. (This may not be a mutual isolation; printed word perception may be informed by the subword level by the presence of a spoken word.) The idea of an independent spoken words processor is consistent with the characterization of speech perception as modular, in the sense of Fodor (1983). It is also consistent with recent data from our laboratory suggesting that the simple detection of the presence or absence of speech in noise is not aided by the concurrent presence of a printed word that corresponds to the spoken word that is embedded in the noise (Frost et al. 1988).

Finally, we suggest that examining the effects of phonological and lexical ambiguity in Serbo-Croatian might provide insights into the difficulties faced by the reader of English. Although there are relatively few words in English that are phonologically bivalent for the skilled reader, there are many more that are phonologically ambiguous for the less skilled reader. The beginning reader of English often comes upon words that are, for him/her, of uncertain pronunciation. The reader may, therefore, decode a string into more than one possible phonological realization that maps onto more than one word in his/her lexicon. The ambiguity may sometimes be resolved by using contextual (i.e., semantic) information but if the phonological information available to the reader is very reduced, even contextual information may not be enough. In either case, the process of lexical access for the less-skilled reader may be similar to the process followed by the skilled reader for phonologically bivalent words. Thus, the phonologically bivalent materials provides the researcher with a way to model natural reading for the unskilled reader by means of the skilled reader. One advantage of the present paradigm is that phonological bivalence for the skilled reader can be introduced and controlled experimentally.

REFERENCES

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 340-357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word frequency effects in the pronunciation task: lexical access and/or production? *Journal of Memory and Language, 24*, 84-106.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation time in reading. *Journal of Memory and Language, 27*, 429-446.

Feldman, L. B. (1987). Phonological and morphological analysis by skilled readers of Serbo-Croatian. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production*. London: Academic Press.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 288-298.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research, 45*, 55-72.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by a simultaneous presentation of print? *Journal of Memory and Language, 27*, 741-755.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17*, 302-311.

Glucksberg, S., Kreuz, R., & Rho, S. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 323-333

Hanson, V. L. (1981). Processing of written and spoken words: Evidence for common coding. *Memory & Cognition, 9*, 93-100.

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 14*, 265-274.

Jastrembsky, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13*, 579-595.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition, 11*, 621-630.

Kroll, J. F., & Schweickert, J. M. (1978). Syntactic disambiguation of homographs. Paper presented at the Nineteenth Annual Meeting of the Psychonomic Society, San Antonio, Texas.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech, 21*, 142-165.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition, 8*, 415-423.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetic word perception. *Journal of Memory and Language, 28*, 237-254.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

Morton, J. (1981). The status of information processing models of language. In H. Longuet-Higgins, J. Lyons, & D. Broadbent (Eds.), *The Psychological Mechanisms of Language*. London: The Royal Society and British Association.

Neill, W. T., Hilliard, D. V., & Cooper, E. (1988). The detection of lexical ambiguity: Evidence for context-sensitive parallel access. *Journal of Memory and Language, 27*, 279-287.

Onifer, W. & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9*, 225-236.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 779-790.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645-657.

Schroeder, M.R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 43*, 1735-1736.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowsky, M. (1982). Automatic access of the meaning of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537.

Simpson, G.B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior, 20*, 120-136.

Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin, 96*, 316-340.

Simpson, G. B., & Burgess. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 28-39.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition, 8*, 513-520.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropsychology, and linguistics*. London, Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language, 26*, 188-208.

FOOTNOTES

[*] *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 569-580. (1990).

[†] Now at the Department of Psychology, The Hebrew University, Jerusalem, Israel.

[††] Also State University of New York at Albany.

[†††] Also University of Connecticut, Storrs.

[1] In general, effects of ambiguity tend to be greater for words than for pseudowords.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 101-109

Morphological Relationships Revealed Through the Repetition Priming Task[*]

Laurie Beth Feldman[†]

INTRODUCTION

Attempts to study morphology experimentally are not new. One methodology for exploring a speakers' knowledge about morphological relationships among words asks subjects directly to judge which words have internal morphological structure. Variants of the method include asking whether a complex word "comes from" a certain base or alternatively, whether pairs of words share some morphological component (e.g., Derwing, 1976, 1986). A second methodology looks for improved performance (i.e., perceptual threshold sensitivity or recognition fluency) over successive presentations of a base morpheme. One task using this methodology employs a variation of the lexical decision task known as repetition priming. Here, morphological variants of a base word are presented and the pattern of facilitation among related forms is interpreted to reflect, at least in part, how those forms are organized in the lexicon of the user.

In the repetition priming procedure (Forbach, Stanners, & Hochhaus, 1974; Scarborough, Cortese, & Scarborough, 1977; Stanners, Neiser, Hernon, & Hall, 1979), each word and pseudoword is presented twice for a lexical decision judgment. The reduction in decision latency relative to a first presentation, that is, the facilitation due to repetition, is measured. (The first presentation of the item is the "prime." The second presentation is the "target.") For facilitation to occur it is not necessary that the identical word be repeated as prime and target. Generally, morphologically related words including inflections and derivations also reduce target lexical decision latencies—sometimes as fully as an identical repetition. This finding invites an interpretation of repetition priming effects as abstract and lexical in origin (Fowler, Napps, & Feldman, 1985: Monsell, 1985; Stanners et al., 1979), although critics of the approach emphasize the perceptual analysis of a specific visual pattern or the memory trace thereof (e.g., Feustel, Shiffrin, & Salasoo, 1983; Jacoby & Brooks, 1984).

Serbo-Croatian, the dominant language of Yugoslavia, has some special properties that have been exploited in the investigation of morphological organization in the internal lexicon. First, numerous inflected and derived variants are used productively in Serbo-Croatian. Consequently, large families of morphologically related words exist. Second, as in English word formation is complex in that there is no unique relation between the form of an affix and its function (compare DISCUSS - DISCUSSION with PROFESS - PROFESSION). Third, Serbo-Croatian is transcribed in two different alphabets, Roman and Cyrillic, and most characters are unique to their respective alphabet. Children learn both alphabets in elementary school and the abundance of printed material in both alphabets guarantees that competence in both alphabets is maintained by adults. As a result, most words in Serbo-Croatian have printed forms in the two alphabets that are visually quite distinct and equally familiar to the skilled reader.

As noted above, one interpretation of repetition priming effects is that the facilitation due to repetition is principally based on the retrieval of information from specific prior episodes or from the repeated perceptual identification of the same pattern in a similar format and context (Johnston, Dark & Jacoby, 1985). Moreover, it has been argued that the "visual characteristics of the display and the configuration of letters in the item ...are probably preserved between successive

---

Terry Lombardi and Terri Ullery ran subjects and Elaine Stotko helped to select and construct materials and I thank them. This work was supported by funds from the University of Delaware. Preparation of the manuscript was supported by NICHD Grant HD-01994 to Haskins Laboratories.

presentations of a letter string." (Feustel et al., 1983: p. 344). In an experiment designed to probe the contribution of visual similarity between prime and target (Feldman & Moskovljević, 1987, Expt 1), morphologically related prime-target pairs were inflected case forms of masculine and feminine nouns in Serbo-Croatian. Primes were printed in either Roman or Cyrillic alphabet characters and targets were always printed in Roman. In that experiment, significant and numerically equivalent facilitation occurred in both the alphabet- preserved (Roman-Roman) and the alphabet-alternating (Cyrillic-Roman) conditions. Recognition latencies to dative case targets were 90 ms faster when either a Cyrillic or a Roman transcription of the same word occurred approximately ten items earlier in the list. This outcome suggests that the basis of facilitation in the repetition priming paradigm lies not with surface attributes that are repeated over multiple instances, such as visual similarity of prime and target, but with some more abstract relationship. This interpretation is consistent with the work of Kirsner and colleagues (Kirsner & Dunn, 1985; Kirsner, Dunn, & Standen, 1987).

In another experiment conducted with Serbo-Croatian materials (Feldman & Moskovljević, 1987, Expt 2), the role of orthographic and phonological similarity between prime and target was investigated by comparing facilitation with real derivations to facilitation with an unrelated monomorphemic word whose structure suggested that it was a derived form. For example, in the triad *KORA*, *KORICA* and *KORAK*, the first member is a nominative form for the word meaning "step" and the second is its diminutive derivation. They are orthographically and phonologically similar but morphologically unrelated to the last member of the triad, which is a pseudodiminutive meaning "crust." Pseudodiminutives are actual monomorphemic words, with a nondiminutive meaning that is unrelated to the target, in which the initial portion is identical to the full target word and the final portion is identical to a diminutive suffix. In that study, significant facilitation occurred for words only when prime and target were morphologically related. Structural similarity of the initial and final portion of prime-target pairs was not sufficient to produce even partial facilitation among word pairs. Pseudodiminutive primes produced no facilitation (1 ms) for structurally similar but morphologically unrelated targets. On the basis of these two experiments it may be concluded that the effects obtained in the repetition priming task cannot be attributed simply to repetition of the same alphabetic (visual) pattern or to repetition of the same orthographic and phonological form. The pattern of facilitation requires the presence of a base morpheme that is repeated across successive presentations. Some theorists have interpreted this outcome as evidence that morpheme units are represented lexically whereas others view it as revealing activation among word forms formed from the same base morpheme.

The present report describes a series of experiments designed to probe the organization of morphologically related words in the mental lexicon of an English speaker. It represents work conducted in collaboration with Carol Fowler and with Elaine Stotko in which the repetition priming procedure of Stanners et al. (1979) was used to investigate constraints on morphological organization. Here, prime and target are morphologically related and the pattern of facilitation to targets as a function of prime type is examined. First, evidence is provided that inflections and derivations produce the same pattern of facilitation. Next, it is shown that primes are equally effective whether or not they preserve the pronunciation and spelling of the base. Collectively, these results suggest that an interpretation of repetition priming based on multiple activations of a lexical entry composed of a base morpheme is inadequate because of the difficulty in forming these variants from the base. In later studies, consideration is given to whether or not the transparency of morphological structure in morphologically complex words is sensitive to the familiarity of the word or to the semantic overlap of prime and target. Finally, evidence that the pattern of facilitation observed with morphological relatives as primes is distinct from that observed with other kinds of semantic relationships is provided.

Inflections and derivations compared

Inflections and derivations differ with respect to the productivity of rules and the predictability of their meaning from a semantic analysis of their constituents. Whereas inflections rarely produce new shades of meaning, derivations are much less constrained semantically and often change meaning once formed (compare TERRIFIC with TERROR). (Aronoff, 1976). Moreover, because letter length of the plausible suffixes varies, derivations tend to be less similar physically to their bases than are inflections, at least in English (Fowler et. al., 1985). Consequently, on linguistic grounds, inflectional and derivational primes might be expected to perform differently. These two types of morphological primes have been compared in the repetition priming task. In one study (Fowler et al., 1985, Expt 2a), triads consisted of uninflected verbal targets, inflected forms with-S and inflected forms with-ED. In a second study (Fowler et. al., 1983, Expt 2b), triads consisted of uninflected verbal targets and two neutral[1] affixed forms (Chomsky & Halle, 1968) selected from the set -MENT, -LESS, -ER, -LY, -NESS, -ABLE, and -FUL. List construction was similar in both studies. Three versions of each test order containing 200 items were created and presented for a lexical decision judgment. Each list included forty-eight word targets and forty-eight orthographically legal but meaningless pseudoword targets. Each pseudoword was matched in length to a word. Targets were preceded forty-eight items earlier in the list by a prime. Targets appeared in the same serial position across all three test orders but the prime varied. Fillers were introduced to maintain necessary lags between prime and target. A subject viewed one test order. A subject saw, therefore, each morpheme twice, once as a prime and once as a target. All subjects viewed three types of primes, either base and S and ED inflections, or base and two different derivations and across test orders each target occurred with each type of prime.

Errors and extreme reaction times (greater than 2000 ms or more than 2.5 SD from the subject's or item's mean) were excluded from the analyses. In addition, when a subject responded incorrectly to the prime, both prime and target responses were eliminated from the analyses. Table 1 summarizes the mean recognition latencies for targets with inflectional and derivational primes.

A one-way analysis of variance with latencies to targets with No Prime, Identity, and Morpheme primes as levels of the factor Condition revealed significant facilitation in both the Identity and Inflectional prime conditions and no significant difference between those two conditions. The analysis of Identity and Derivational forms produced a similar pattern. Facilitation to a base form by a derivational relative was as strong as facilitation by an inflectional relative.

That the pattern of facilitation for inflected and derived primes was the same suggests that the basis of repetition priming among forms cannot be a common lexical entry from which all morphological variants are generated by rule. While it is plausible that facilitation among inflected forms and bases occurs because the same lexical entry is activated twice, and it is reasonable to assume that the base form constitutes the lexical entry from which inflectional variants are generated by rule, based on issues of productivity and semantic predictability the base form plus rule account cannot apply to derived words. Alternatively, it can be proposed that facilitation derives from activation among words (as contrasted with a base morpheme) that are members of a family of morphologically related words and that these relationships are stronger than the relationships among semantically related but morphologically-unrelated words. This interpretation was supported in a subsequent study (Fowler et al., 1985, Expt 3) that examined repetition priming among words that cannot be generated from each other either by simple rules of inflection or derivation.

Table 1. *Mean decision latency (ms) to targets preceded by Identity, Inflectional and Derivational primes.*

	PRIME				
	No Prime	Identity	Inflectional	Derivational	
		manage	manages	management	prime
	manage	manage	manage	manage	target
WORDS					
	611	510	533		
	585	543		538	
PSEUDOWORDS					
	643	627	645		
	715	717		730	

From Fowler, Napps and Feldman, 1985, Expt 2.

Do alternations in spelling and pronunciation obscure morphological relationships?

It is evident that morphological relationships among regular morphological variants are discovered by skilled readers. Are morphological relationships between words that undergo alternations to the base morpheme similarly handled by the skilled reader? Using the method described above, subjects performed a lexical decision judgment to primes and targets that were morphologically related. In this study (Fowler et al., 1985, Expt 3), morphological relatives consisted of triads composed of one base form and two affixed forms one of which preserved the spelling and pronunciation of the base and one of which did not. In the Sound Only set, one affixed form preserved both sound and spelling and the other preserved only the spelling. An example is HEAL, HEALER, HEALTH. In the Sound and Spelling Set, one affixed word had the same spelling and pronunciation of the base component as the unaffixed word while the other did not. An example is CLEAR, CLEARLY, CLARIFY. Because no unambiguous difference in facilitation for inflected and derived forms has been observed with this technique in the previous study (but see Stanners et al., 1979), sets included both inflections and derivations. Moreover, neutrally and nonneutrally affixed forms were selected. Unaffixed words in the Sound only and the Sound and Spelling lists were matched on length and frequency as were the affixed forms. Phonological overlap between the unaffixed and affixed forms was matched across lists by counting each vowel, consonant or stress change as a change. As in earlier experiments, the lists also contained pseudowords derived from real words by changing one or two letters in other words.

Three test orders were created in a manner analogous to that described in the previous experiment. Each base morpheme appeared once in a prime and once in a target. The test orders differed with respect to prime (Identity, Morpheme-No Change, Morpheme-Change), and the serial position of each target was constant over lists. Targets were always the unaffixed form. Half of the targets were from the Sound only set and half from the Sound and Spelling set and each type of prime (Identity, No Change, Change) was equally represented. In all, there were forty-eight word targets and forty-eight pseudoword targets, and lists were constructed so that the interval between prime and target was forty-eight items. A subject viewed one test order. All subjects viewed all three types of primes on different items and across test orders each target was preceded by each type of prime.

Errors and extreme reaction times were deleted from all analyses by the procedure described above. Results of an analysis of variance revealed significant facilitation when a morphologically related prime preceded the target. Table 2 summarizes decision times for word and pseudoword targets preceded by primes that undergo change and primes that do not change (Fowler et al., 1985, Expt 3). The effect of stimulus set (Sound only/Sound and Spelling) was significant but did not interact with type of prime.

Table 2. *Mean decision latency (ms) to targets preceded by Identity, No Change, and Changed Morphological primes.*

		PRIME			
	No Prime	Identity	No Change	Change	
		clear	clearly	clarify	prime
	clear	clear	clear	clear	target
WORDS					
	623	558	575	584	
PSEUDOWORDS					
	760	748	758	746	

From Fowler, Napps, & Feldman, 1985, Expt.3.

That is, the same pattern of facilitation was observed for both sets of stimulus materials. Most important, among the prime conditions, Identity differed from No Change and Change primes but the latter two conditions did not differ one from the other. That is, the effect of sharing or not sharing spelling or pronunciation with the base (viz., target reaction times were 575 and 584, respectively) was not significant. Finally, neutral and nonneutral derivations were compared in order to determine whether the equivalence of inflectional and derivational primes reported earlier (Fowler et al., 1985, Expt 2) is restricted to derivations that preserve the pronunciation of the base morpheme. The outcome mirrored the overall analysis and indicated no difference between No Change (neutral) and Change (nonneutral) derivations with respect to facilitation of an unaffixed target.

In summary, the major finding of Fowler and her colleagues (1985, Expt 2) was that the magnitude of facilitation is not reduced when the orthographic or phonological forms of affixed primes and unaffixed targets do not fully overlap. Because effects of form overlap were not observed, effects of orthographic and phonological similarity could not be distinguished. Evidently, regularly affixed words, and words that undergo alternations, are equally effective primes for their morphological relatives in the repetition priming task. This outcome supports the claim that the facilitation observed in the repetition priming task cannot be due to multiple activations of the same base morpheme where all morphological relatives are derived from the base by rules of inflection or derivation. Alternatively, the claim that facilitation derives from activation among members of a family of morphologically related words is supported. What are the constraints on the activation between word forms? In the next study (Feldman & Stotko, unpublished data a), we asked whether the familiarity of a morphologically complex word is critical to recognition of its base morpheme.

Are less familiar words analyzed more readily with respect to morphemic structure than more familiar words?

Prior to conducting a repetition priming procedure, a group of thirty-five subjects rated the familiarity of a list of words in order to obtain a measure of familiarity. Subjects were instructed to use the full range of scores, from 1 to 7 of the scale. Examples of very familiar and very unfamiliar words were provided. The list contained clusters of words that differed in frequency and shared a base morpheme although members of a cluster were distributed throughout the list. Based on the familiarity ratings, a collection of twenty-seven word triads was selected. Each triad included a target noun and two morphological relatives, and all members were known to all raters. One member of each triad was more familiar than the target (High familiarity prime) and the other was less familiar than the target (Low familiarity prime). Both morphological relatives were derivations of the base noun. For example, the high and low familiarity primes for the target MULTIPLE were MULTIPLY and MULTITUDE. In the experimental list, for half the triads, the base form served as target and for the other half a derivation was the target.

The lists also contained pseudowords that were constructed by changing up to three letters in the root of one of the target words. All substitutions formed legal letter strings. For example, *MACTICAL* was derived from *PRACTICAL* and *DRACIALIZE* from *SPECIALIZE*. The pseudowords were also presented in "morphologically related" high and low familiarity forms. For example, *DRACIAL* and *DRACIALITY*.

For the lexical decision task, three test orders containing 114 items were created. Each included equal numbers of word and pseudoword targets and all three types of primes (viz., Identity, High and Low). There were twenty-seven word targets and twenty seven pseudoword targets, each of which was preceded earlier in the list by its prime. Lists were constructed so that the interval between first and second presentations of a word (or pseudoword) ranged from seven to thirteen items with an average of ten. In order to maintain appropriate lags, six filler items were also included. Words appeared in the same serial position across all three test orders but the prime varied. For example, the word *AUDIBLE* was presented as the target in all three orders but within each test order it was preceded, in the same position, by either *AUDIBLE*, *AUDIENCE* or *AUDITION*.

A subject viewed one test order. Each subject saw, therefore, every morpheme twice, once in a prime and once in a target. All subjects viewed all three types of primes on different items and across test orders each target was preceded by each type of prime. No subject participated in both the familiarity assessment and the lexical decision tasks.

Errors and extreme reaction times (greater than 1250 ms or less than 350 ms) were eliminated from all analyses. In addition, when a subject responded incorrectly to one member of a prime-target pair, both responses were eliminated from subsequent analyses. The error pairing procedure eliminated an additional 3% of all responses. Table 3 summarizes the mean recognition times for words and pseudowords with highly familiar and less familiar primes for the Feldman and Stotko data.

A one-way analysis of variance with latencies to targets with No prime, Identity, High and Low familiarity primes as levels of the factor Condition revealed significant facilitation in the Identity condition and numerically reduced but statistically significant facilitation when either derivation served as a prime. Most important, there was no significant difference between primes that were of higher and lower familiarity than the target.

Evidently, the relative familiarity of primes does not influence the magnitude of facilitation observed in the present task. High and Low familiarity primes produced statistically equivalent and numerically similar measures of facilitation despite differences in recognition latency to the primes themselves (583 High vs 694 Low). The differences between latencies to primes can be interpreted as evidence that the primes did, in fact, differ on familiarity. Nevertheless, the present outcome provides no evidence that the morphological structure of less familiar words is considered more readily than that of more familiar words. In this respect, morphological analysis contrasts with the phonological analysis of a word because it is not sensitive to a word's frequency of use (e.g., Seidenberg, 1985). It should be mentioned that because of the constraints of the present task, only correct responses provide interpretable data. Consequently, no extremely unfamiliar words were included. It is possible that effects might still emerge with another task and a more expanded range of familiarity. Another possibility is that the pattern of facilitation is determined by the semantic similarity of prime and target. This outcome was reported for semantically related pairs (Fischler, 1977) but has not been examined for morphological relatives. The hypothesis that semantic similarity of prime and target influences the pattern of facilitation observed in repetition priming was explored in the next study (Feldman & Stotko, unpublished data b).

Does semantic similarity enhance the recognition of morphological relationships?

In order to establish a measure of semantic overlap, a group of forty subjects judged which of two word pairs was closer semantically. The list contained thirty-eight pairs of pairs such as *CREATE / CREATURE* and *CREATE / CREATION* and subjects were instructed to indicate which of the pairs was more similar in meaning. Based on the responses of the initial group of subjects, a collection of twenty-seven triads was formed. Each included a semantically close and a semantically far morphological relative. All relatives were derivations. Examples of triads with Identity, Close and Far primes are ACTOR/ ACTION/ ACTUARY; PEDAL/PEDESTRIAN/ PEDESTAL; CREATE/CREATION/CREATURE. Pseudowords were constructed in the manner described above. Note that the distinction between close and far is meaningless here. Examples of pseudoword triads are SICTOR/SICTION/SICTUARY; CHADEL/ CHADESTRIAN/CHADESTAL; BREATE/ BREATION/BREATURE.

Table 3. *Mean decision latency (ms) to targets preceded by Identity, High and Low Familiarity primes.*

	PRIME				
	No Prime	Identity	High	Low	
		multiple	multiply	multitude	prime
	multiple	multiple	multiple	multiple	target
WORDS	609	515	585	590	
PSEUDOWORDS	690	653	689	686	

From Feldman & Stotko, unpublished data a.

For the lexical decision task, three test orders each containing 114 items were created according to the constraints of lag and the ratio of words to pseudowords adopted in the previous study. Each subject viewed all types of primes and across test orders, each target was preceded by its Identity, Close and Far semantic prime. No subject participated in both the semantic judgment and lexical decision tasks.

Errors and extreme response times (greater than 1100 or less than 350 ms) were eliminated from all analyses. In addition, responses were deleted according to the error pairing procedure described above. By this procedure, an additional 2% of all responses were eliminated. Table 4 summarizes mean recognition times over subjects for the Feldman and Stotko data.

A one-way analysis of variance with latencies to targets in the No Prime, Identity, Close and Far semantic prime conditions showed significant facilitation in the Identity condition and dramatically diminished facilitation for both semantically Close and Far derivations. The reduced facilitation in the semantically Close and Far conditions relative to the Identity prime condition was unexpected. Central to the current focus of discussion, there was no significant difference in the magnitude of facilitation observed with primes that varied in their degree of semantic overlap with the target. This outcome is not surprising if the facilitation observed in the repetition priming paradigm is believed to be distinct from that observed when prime and target are semantically related but morphologically distinct. This claim was tested in the fifth and final study to be described.

It has been demonstrated that repetition priming persists over longer lags than semantic or associative priming. Whereas effects of repetition priming occur over intervals of two days (Scarborough et al., 1977) and lags of 48 intervening items (Fowler et al., 1985), effects of semantic priming are not evident at lags of two items (deGroot, 1983; Dannenbring & Briand, 1982; Davelaar & Coltheart, 1975). The distinction in the duration of facilitation has led some researchers (e.g., Henderson, Wallis & Knight, 1983; Den Heyer, Goring & Dannenbring, 1985) to posit distinct underlying principles of lexical organization for morphological relationships and other semantic relationships. A similar conclusion derives from a comparison of neural events during different experimental conditions. In particular, event-related potentials, transient perturbations to the electroencephalogram after the onset of an external event, vary under conditions of semantic and repetition priming (Rugg, 1987). Finally, the semantic similarity of prime and target correlates positively with the magnitude of facilitation in a semantic priming task (Fischler, 1977) although, as reported above (Feldman & Stotko, unpublished data b), no effect of semantic overlap has been observed in repetition priming. In order to probe further the claim that morphological priming is distinct from other types of semantic priming, a study was conducted (Fowler & Napps, unpublished data) in which target words were preceded by highly associated antonyms or by themselves. If subjects are able to anticipate upcoming target items, or if the basis of facilitation in repetition priming is the presence of a semantic relationship regardless of morphological considerations, then antonyms should reveal facilitation in the present experimental setting. Alternatively, if a morphological relationship is a precondition for facilitation at these lags then no facilitation by antonyms is anticipated.

Subjects generated associations to a set of target words, items for which at least 90% of subjects provided the same association (an antonym) were selected as antonym primes. There were 54 word targets and an equal number of pseudoword targets. Lists were constructed to include both words and pseudowords and Identity and Antonymprimes (for words). The interval between prime and target was 36 items. Words appeared in the same serial position across test orders but the prime that preceded it varied. For example, in one test order, *COLD* was preceded by *HOT* and in the other it was preceded by its identical form, *COLD*.

By analogy to the experiments described above, a subject viewed one test order that included both Identity and Antonym primes and across test orders each target was preceded by both types of prime. No subject participated in both the association and the lexical decision phases of the experiment.

Errors and extreme scores were eliminated according to the criteria described in Fowler et al., (1985, Expt 1). In addition, responses were eliminated according to the error pairing procedure described above. Results are summarized in Table 5.

A one-way analysis of variance with latencies to targets with No Prime, Identity and Antonym primes revealed statistically significant facilitation in the Identity condition and no facilitation in the antonym condition. The presentation of a highly associated and semantically related prime did not facilitate recognition of the target.

Table 4. *Mean decision latency (ms) to targets preceded by Identity, Close and Far Semantic primes.*

		PRIME			
	No Prime	Identity	Close	Far	
		create create	creation create	creature create	prime target
WORDS	543	450	507	513	
PSEUDOWORDS	615	590	600	593	

From Feldman & Stotko, unpublished data b.

Table 5. *Mean decision latency (ms) to targets preceded by Identity and Antonym primes.*

		PRIME		
	No Prime	Identity	Antonym	
		cold	hot	prime
	cold	cold	cold	target
WORDS	625	581	628	
PSEUDOWORDS	696	691	697	

From Fowler & Napps, unpublished data.

As noted above, this outcome is not anticipated if the basis of facilitation in repetition priming is purely semantic in origin or if subjects can anticipate upcoming targets. Conversely, if the basis of facilitation in the repetition priming task exploits the relationship among words that share a base morpheme, then this result is readily interpretable. In conclusion, it is morphological relationships that govern facilitation in this task.

DISCUSSION

In a series of studies employing the repetition priming paradigm, significant facilitation among word forms was observed for a) regular inflected and derived forms and b) forms that undergo alternations in pronunciation and spelling. In addition, the magnitude of facilitation to target latencies was equivalent for inflectionally and derivationally related primes. Facilitation to target words in this task was interpreted as evidence that these words share a lexical entry. This entry must fully specify all of the morphologically related forms, however, because many variants cannot be generated from the base morpheme by general rules. Finally, relatedness within an entry is not a function of familiarity of word forms nor of the semantic similarity of forms because neither factor influenced the magnitude of facilitation. By contrast, experimental measures of relatedness may be sensitive to the formal overlap or the regularity between forms (Fowler et al., 1985).

Repetition priming and semantic or associative priming both reveal principles of organization for lexical knowledge. They contrast with respect to longevity of facilitation and to the role of semantic similarity of prime to target. Semantic facilitation does not occur at long lags whereas repetition priming does. In addition, neither the factors of semantic overlap of morphologically related pairs nor strong semantic relationships without morphological relationships (viz., antonym pairs) figure prominently in the results observed with the repetition priming paradigm although they are evident in semantic priming. These differences support the claim that the principle of lexical organization for morphological relationships and other semantic relationships are distinct.

In conclusion, the outcome of a series of experimental studies suggests that lexical entries for sets of morphologically related words include more words than simply those variants that can be generated from the base morpheme by a rule: Neither the spelling nor the pronunciation need be preserved. In addition, the meaning of morphologically complex forms need not be predictable from that of the base and its affixes. In summary, all morphological relationships are captured in the mental lexicon of the skilled reader.

REFERENCES

Aronoff, M. (1976). *Word formation in generative grammar* (2nd printing, 1981). Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology, 36*, 435-444.

Davelaar, E., & Coltheart, M. (1975). Effect of interpolated items on the association effect in lexical decision tasks. *Bulletin of the Psychonomic Society, 6*, 269-272.

deGroot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*, 417-436.

Den Heyer, K., Goring, A., & Dannenbring, G. L. (1985). Semantic priming and word repetition: The two effects are additive. *Journal of Memory and Language, 24*, 699-716.

Derwing, B. (1976). Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics, 21*, 38-66.

Derwing, B. L., & Nearey, T. M. (1986). Experimental phonology at the University of Alberta. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 187-209). Orlando: Academic Press.

Feldman, L. B., & Moskovljević, J. (1987). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 573-581.

Feldman, L. B., & Stotko, E. (unpublished data a).

Feldman, L. B., & Stotko, E. (unpublished data b).

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112*, 309-346.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition, 5*, 335-339.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition, 2*, 337-339.

Fowler, C. A., & Napps, S. (unpublished data).

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241-255.

Henderson, L., Wallis, J., & Knight, D. (1983). Morphemic structure and lexical access. In H. Bouma & D. Bouhuis (Eds.), *Attention and performance*. London: Erlbaum.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. *The psychology of learning and motivation, 13*, 1-46.

Johnston, W., Dark, V., & Jacoby, L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 3-11.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 547-565). Hillsdale,NJ: Erlbaum.

Kirsner, K., Dunn, J., & Standen, P. (1987). Record-based word recognition. In M. Coltheart (Ed.), *Attention and performance XII The psychology of reading* (pp. 147-167). Hove: Erlbaum.

Monsell, S. (1985). Repetition and the lexicon. In W. Ellis (Ed.), *Progress in the psychology of language* (pp. 147-195). London: Erlbaum.

Rugg, M. D. (1987). Dissociation of semantic priming, word association and non-word repetition effects by event-related potentials. *Quarterly Journal of Experimental Psychology, 39A*, 123-148.

Scarborough, D. L., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 1-17.

Seidenberg, M. S. (1985). The time cause of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Stanners, R. F., Neiser, J. J., & Painton, S. (1979). Memory representation for prefixed words. *Journal of Verbal Learning and Verbal Behavior, 18*, 733-743.

FOOTNOTES

[*] To appear in M. Noonan, P. Downing, & S. Lima (Eds.), *Literacy and linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Co.

[†] Also State University of New York at Albany.

[1] Two types of suffixes exist in English (Chomsky & Halle 1968). Neutral affixes include inflections and some derivations and these affixes do not affect the pronunciation of the stem to which they are attached. By contrast, nonneutral affixes do affect the pronunciation of the stem. According to Chomsky and Halle's theory, neutral affixes are separated from the stem morpheme by a word boundary whereas nonneutral affixes are separated by a morpheme boundary and phonological rules apply over morpheme but not word boundaries.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 110-129

Listener-talker Attunements in Speech*

Carol A. Fowler†

The phonetic information that listeners extract from the acoustic speech signal does not correspond to features of the signal that are obvious in visible displays of speech. Listeners can report hearing sequences of the phonetic segments (consonants and vowels) that alphabetic orthographies spell, but phonetic segments are not evident in displays of acoustic speech signals. Whereas acoustic segments can be identified reliably in these displays, they do not correspond to phonetic segments (e.g., Fant & Lindblom, 1961): there are many more acoustic than phonetic segments; and each acoustic segment provides information about more than one phonetic segment; each phonetic segment affects, and hence is signaled by, a variable number of acoustic segments.

We understand the reasons why acoustic segments relate in so complex a way to phonetic segments. Talkers coarticulate gestures for different phonetic segments (that is, they produce them in overlapping time frames). Therefore, components of neighboring phonetic segments are, at best, interleaved, and some researchers suggest that they are even changed (in being assimilated to their neighbors; e.g., Daniloff & Hammarberg, 1973) or, worse yet, "distorted" (Ohala,1981) or "smashed" and "rubbed together" (like raw eggs sent through a wringer; Hockett, 1955) or "encoded" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) in articulation and therefore in the acoustic speech signal.

Phonetic perception and visibly displayed acoustic signals fail to correspond in another way as well. Early research on perception of speech uncovered a variety of "cues" important for the perception of phonetic information. For example, the second-formant transition[1] in CV syllables provides information for (is a "cue" for) consonantal place of articulation.[2] Synthetic syllables can be created in which the second-formant transition provides the only information in the signal for place of articulation, and listeners will correctly report the consonant whose place information is specified in that way. Figure 1 shows the different transitions for /d/ in /di/ and /du/ in synthetic syllables identified reliably and correctly by listeners. In these syllables, the very different second formant transitions provide the only information for stop place. Yet listeners hear both synthetic syllables as having syllable-initial /d/s; moreover, they hear the /d/s as indistinguishable (Liberman et al. 1967). When the transitions are removed from their contexts and presented in isolation to listeners, each sounds like the frequency glide it resembles in a visual display, and neither sounds like a "d." Here, then is another kind of disagreement between obvious characteristics of the acoustic signal and the listener's percept. The phonetic segments do not sound, phenomenally, like a composition of the the visibly obvious acoustic fragments that nonetheless do somehow support phonetic perception.

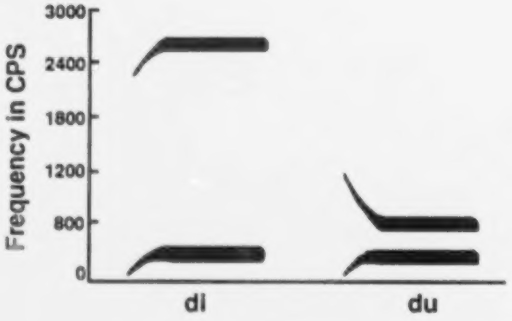


Figure 1. Synthetic tokens of syllable "di" and "du" in which the second formant transitions specify the consonant.

Preparation of this chapter was supported by a fellowship from the John Simon Guggenheim Foundation and by NICHD Grant HD-01994 and NINCDS Grant NS-13617 to Haskins Laboratories.

An implication of findings such as these for some theorists is that the phonetic percept is generated "in the head" based only in part on the acoustic input, and in part on stored knowledge of phonetic segments or syllables and the acoustic "cues" that hint at their occurrence in speech (e.g., Hammarberg, 1976; 1982; Massaro, 1987). In Massaro's theory, for example, achieving a percept involves choosing a syllable-template in memory that best fits the collection of incoming cues. There is a different template for /di/ and /du/, so different cues can signal the initial consonant they share. Missing from this approach, however, is an explanation for how the stored knowledge got there (how it was learned) if the acoustic signal and the phonetic percept *must* be mediated by stored knowledge. Missing, too, is an explanation why the second-formant transitions for /di/ and /du/, in their syllabic contexts, sound alike. Why do we spell the initial sound of these syllables with the same letter? Why is there no language in the world in which these acoustically different segments are different phonemes of the language? More fundamentally, why can't we tell the sounds apart? In short, what makes a /d/ a /d/?

The motor theorists, who uncovered the striking /di/-/du/ result, noticed that the phonetic percept does correspond to the talker's *production* of /d/. /d/ is invariantly produced by a gesture of the tongue tip, which temporarily seals off the vocal tract behind the teeth. The mouth is shut during the invariant part of the gesture, however, and so the signal is largely silent then. The formant transitions differ in /di/ and /du/, because they reflect the opening movement away from the closure to a vocal-tract shape that varies with the vowel. Listeners somehow extract the invariant articulatory gesture for the /d/ from a context-sensitive acoustic signal. The motor theorists, like most other theorists, see coarticulation as a process that prevents intended phonetic segments from being realized nondestructively in the vocal tract. Accordingly, for them, as for the theorists cited above, phonetic perception must be mediated (Liberman et al., 1967), but, here, by means of a decoding mechanism that recovers phonetic gestures from the acoustic signal. The motor theorists realized that if coarticulation ensures a less-than-fully-informative acoustic signal, the decoding processes cannot be learned and so must be innate, part of the human's biological specialization for language.

A second consequence of the failure of the acoustic signal to specify perceived phonetic gestures fully is that the decoding mechanism must operate by guesswork. Liberman and his colleagues (Liberman et al., 1967; Liberman & Mattingly, 1985) propose that it works by a process of "analysis by synthesis" (Stevens, 1960; Halle & Stevens, 1962). In its most recent version (Liberman & Mattingly, 1985), the decoding mechanism uses the acoustic speech signal to generate a few candidate phonetic gestures that may have produced it. These are evaluated against the signal by means of an "internal, innately specified vocal-tract synthesizer" (p. 26) that knows the alterations to the gestures that coarticulation effects at the articulatory surface and knows the acoustic consequences of those alterations. According to Liberman and Mattingly, "the module [the innate perceiving mechanism] has then merely to determine which (if any) of the small number of gestures that might have been initiated at a particular instant could, in combination with gestures already in progress, account for the signal" (p. 27).

I find two shortcomings in this approach. One derives from the fact that coarticulatory influences of a gesture operate bidirectionally so that a gesture for a phonetic segment coarticulates with gestures both for preceding and for following segments. An innate vocal-tract synthesizer can determine the probable coarticulatory influences of a gesture for an already-identified phonetic segment on a segment currently being evaluated, but it cannot be expected to work out the coarticulatory influences of gestures for segments that it has not yet identified. In short, I am not sure that the decoder can work as described. A related criticism is that the innate mechanism will not work reliably if it has to guess. Talkers produce phonetic segments at rapid rates thanks to coarticulatory overlap (Miller, Grosjean, and Lomanto [1984] report an average of 4.97 syllables per second, and so between about 10 and 15 phonemes per second in stretches of speech uninterrupted by pauses.). Listeners cannot afford to guess wrong very often if they are to keep up. Yet if the "guesses" are not wrong, then they are not guesses. If they are not guesses, then the acoustic signal must provide a better source of information about phonetic gestures than its foregoing characterizations in terms of acoustic segments and cues would suggest.

Some investigators have begun to look for new perspectives on the acoustic speech signal that may bring its description into better alignment with perceived phonetic properties. Stevens and Blumstein (e.g., 1981) and Blumstein and her colleagues (e.g., Lahiri, Gewirth, & Blumstein,

1984) have looked for invariant information in the acoustic speech signal that may directly specify phonetic features.[3] (See also Kewley-Port, 1983.) Stevens and Blumstein (1978) proposed invariant acoustic information for place of articulation in the integrated, short-time spectra of stop consonants at release.[4] In particular, they proposed that the spectra for bilabial (/b/, /p/), alveolar (/d/, /t/) and velar (/g/, /k/) places of articulation differ in overall shape, with bilabial spectra flat or falling over the frequency range from high to low, alveolar spectra having more energy at higher frequencies and velar spectra having most energy in the middle range.

These prototypical spectra do well in classifying naturally-produced stops of English, particularly in syllable-initial position, by place of articulation (Blumstein & Stevens, 1979). However, they have proven inadequate in two respects. First, extended to other languages (in particular, French and Malayalam; Lahiri, et al., 1984) with a dental place of articulation between the bilabial and alveolar places, the spectra fail to distinguish the bilabial and dental places. Second, listeners do not appear to use the spectra as significant sources of information for place. Both Blumstein, Isaacs, and Mertus (1982) and Walley and Carrell (1983) pitted the invariant spectra against context-sensitive information for place (the starting frequencies of unchanging formant transitions in Blumstein et al. and the second-formant transitions in Walley and Carrell) and found that listeners used the context-sensitive information consistently to identify place.

Recently, Lahiri et al. (1984) have proposed new invariants for stop place that reflect changes in spectral energy over time between stop release and onset of voicing for the next vowel. These invariants properly classify bilabial and alveolar (English) or bilabial and dental (French) stops. Moreover, pitted against context-sensitive information, they determine listeners' place judgments.

An important advantage of the approach of Lahiri et al. over that of Stevens and Blumstein is that it looks for information specified over time, not in a static snapshot of the vocal tract at release. This constitutes an advance, because stop gestures are events that occur in time. The approach can be improved, however. A small problem, perhaps, is that the invariant proposed by Lahiri et al. will not explain the perception of the synthetic /d/s in /di/ and /du/ depicted in Figure 1, because there is no energy in the spectra of either syllable in the ranges where Lahiri et al. find invariant information for /d/. A more important failing, in my view, is that Stevens and Blumstein and Blumstein and her colleagues are in fact looking for invariant information that specifies the static phonetic features of Jakobson, Fant, and Halle (1952). However, the structure in the acoustic speech signal is the product of articulatory gestures—linguistically-organized to be sure—but gestures all the same. The phonetic features of Jakobson et al. are attributes of those gestures (Fowler, 1977; Grundt, 1976). (In Weiss' [1969] metaphor referring to genes, they are adjectival properties of the (noun-like) gestures.) Thus, missing from this research program so far is recognition, following the motor theorists, that listeners recover gestures. Invariants will specify the gestures, not merely their attributes. (So, for example, information for a "bilabial" consonant will be information not for the features [+grave] and [+diffuse], but for a closing (and/or release) gesture at the lips. Indeed, that presumably is why the spectral changes over time, proposed as invariants by Lahiri et al., work better than the static spectra of Stevens and Blumstein.[5]

As I have presented them, all of the foregoing approaches to understanding phonetic perception are on the right track in some respects, but not in others. Certainly as language learners acquire knowledge about the component words, syllables and other units of the language, that knowledge can facilitate recovery of information from noisy signals (e.g., Massaro, 1987). Too, possibly in our evolution as speaker/hearers we have evolved specialized neural mechanisms dedicated to recovery of phonetic information from speech signals (Liberman et al., 1967; Liberman & Mattingly, 1985). Evidence for laterality of speech perception (e.g., Kimura, 1961; Studdert-Kennedy & Shankweiler, 1970) favors that view. However, I doubt that we have the stored knowledge or an innate mechanism, if we do, *because* coarticulation transforms, distorts, smears and encodes phonetic segments as produced, thereby rendering the acoustic signal inadequate to specify the talker's intended phonetic message. If that were the case, the needed top-down information could not be acquired and the innate "decoding" mechanism would have to guess. In short, these theories of perception seem to require as a premise that, in fact, the acoustic signal does specify phonetic segments as produced.

To find acoustic information for gestures, we need another way to understand coarticulation and, based on that, another way to look at the acoustic signal as a source of information about gestures. A search for invariants may be on the right track; however, the search should be constrained by the very compelling evidence that listeners recover what talkers do gesturally.

In the following three sections of the chapter, I will develop the approach I am advocating. I will first offer a perspective on coarticulation (see also Fowler, 1983a; Fowler, Rubin, Remez, & Turvey, 1980) as a means of ordering segments serially in a compact, but nondestructive manner by means of articulatory "layering." I will suggest that coarticulation is not different from another kind of layering in speech production, that of segmental and "suprasegmental" (intonational, stress-related) information that has generally not been identified as assimilative transformation, distortion, smearing or encoding.

Next I will review evidence uncovered largely by proponents of the motor theory that listeners do indeed recover articulated phonetic segments. Most of this evidence, looked at one way, provides additional examples of the lack of correspondence between percept and cue demonstrated so strikingly in the example of /di/ and /du/. In other theoretical approaches, the lack of correspondence is seen as evidence that phonetic segments are "internally generated, the creature of some kind of perceptual-cognitive process" (Hammarberg, 1976, p. 355). However, there is another interpretation that I find more plausible, which derives from the observation that this kind of correspondence failure is general to perception. The alternative interpretation is that the correspondence failure reflects the fact that proximal stimuli are not themselves objects of perception, but rather serve as information for perception of the environmental sources of their structure (cf. Gibson, 1966; 1979). Reflected light does not resemble the properties of environmental events it specifies to visual perceivers, but it does serve as a reliable source of information about the events that gave rise to its structure over time. By the same token, we should not expect displays of the acoustic signal to resemble phonetic segments; rather we should expect the signal itself to provide reliable information for those segments, because articulation of the segments produced the structure in the signal. In a speech signal, formant transitions do not provide information for a frequency glide, but rather for a change in the size of cavities of the vocal tract as the jaw closes or opens to realize a gesture for a phonetic segment.

In the third section, I will look at some additional failures of listeners to hear speech in a way that corresponds to the appearance of the acoustic signal, visibly displayed. These new examples will show that listeners hear coarticulated speech, and layered segmental and suprasegmental information more generally, in a way that corresponds closely to the talker's manner of producing it. While it has taken speech researchers a long time to begin to understand coarticulation and suprasegmental layering, listeners have been sensitive to their structure all along. Listeners are remarkably attuned to talkers' behavior in producing speech.

Of course, listeners are not always successful in recovering what talkers do. Mishearings occur sporadically among experienced perceivers (e.g., Garnes & Bond, 1980; Browman, 1980), and sometimes they may even occur systematically (e.g., Ohala , 1974; 1981). However it may be that a theory of speech perception explains perception among experienced users of the language, it must suppose that language learners extract information for articulation. Children must learn to utter the speech of their language community by extracting information for articulation from the speech they hear around them. In Ohala's view (1974; 1981), perceivers (presumably, language learners) may extract wrong information by failing to recognize gestural layering as such.[6] This sometimes promotes sound changes in the language whereby pronunciation of words or classes of words changes in a language community. These diachronic products of perception failures are identifiable in part because they are phonological forms or processes that generate acoustic signals similar, though often exaggerated and stylized, to acoustic products of gestural layering. In a sense they constitute fossilized remains of the fundamental tendency of language users to perceive articulatory gestures. In the final section of the chapter, I will provide some examples of those fossilizations.

Coarticulation and other gestural layering in speech production.

Phonetic segments as produced are composed of gestures. For example, [p] consists of a gesture that creates a complete occlusion of the vocal tract at the lips and then releases it after sufficient air pressure has built up behind the closure. In addition, it includes, at least, a gesture abducting the vocal folds, if they are not already abducted, and one raising the velum (to prevent airflow through the nasal cavity), if it is not already raised. Each gesture may involve one or more articulators. The closing gesture, for example, involves the jaw and the upper and lower lips.

Articulators responsible for the closing gesture are coordinated--that is, functionally coupled--to ensure achievement of the gesture in a variety of contexts.[7] Possibly the different gestures for a phonetic segment are coordinated as well, although the research needed to test that possibility has not been done. Certainly, voiced and voiceless stops require delicate relative timing of the closing and release gestures of the oral cavity with the laryngeal gesture adducting or abducting the vocal folds (Löfqvist & Yoshioka, 1980). For example, a voiceless stop in stressed syllable-initial position includes a glottal opening gesture that nearly coincides with stop release (Löfqvist, 1980).

The gestures for a phonetic segment are not often synchronous, however. For example, the lip rounding gesture for a rounded vowel such as /ɔ/ in "claw" anticipates the jaw-opening gesture for the vowel. In the example, it may begin during the earliest gestures for the word-initial /k/. Likewise lowering of the velum for a nasal consonant occurs early relative to the closing gesture for the nasal in the oral cavity. The gestures for a segment do not end synchronously either. The nonsynchrony of the gestures for a given segment and the consequent temporal (and sometimes, when an articulator is shared by two gestures, spatial) overlap among gestures for different segments is known as "coarticulation."

Coarticulation is sometimes interpreted as an assimilatory process of feature sharing (Daniloff & Hammarberg, 1973), so that, if, for example, the rounding gesture for /ɔ/ in "claw" extends into the domain of /k/ and /l/, the consonants are said to have acquired the phonetic feature [+ rounding]. This interpretation does not fit the facts, however. It does not fit the observation that rounding may extend only part way into a consonantal segment in the neighborhood of a rounded vowel. In linguistic theory, however, segments cannot become, say, one-third rounded. It does not fit the perceptual facts either. As the research shows that I review under "Perceiving cohesion and separation of gestures for coarticulating phonetic segments," listeners ascribe the coarticulatory effects of a phonetic segment to the coarticulating segment; they do not hear the affected segments as assimilated to their contexts. Finally, the account of coarticulation as feature spreading probably obscures the reasons why talkers coarticulate. Most likely, they do so for the same reasons that typists coarticulate finger movements (Gentner, 1983). It is efficient to do so (as Liberman et al., 1967 point out), and it does not impair communication of the linguistic message as long as the serial ordering and identity of letters in typing and of phonetic segments in speaking is preserved.

At worst, coarticulation is seen as a destructive process that distorts or smears concurrently produced gestures so that phonetic segments become thoroughly encoded in the speech signal. These descriptions do (in one way of looking at articulation) appear to fit the articulatory facts and they even may be consistent with the reasons why talkers coarticulate. They do not fit the perceptual facts well, however. The segments are not distorted or smeared for listeners, and I see no other way to explain that than to suppose that they are not smeared or distorted in articulation, or, therefore, in the acoustic speech signal. It is a job for speech researchers to find out what talkers do to preserve phonetic-segmental integrity in production and what information the acoustic speech signal provides listeners.

I suggest (e.g., Fowler, 1983ab; Fowler et al., 1980) that the overlap of gestures in coarticulation is a nondestructive *layering* of gestures--most simply of gestures that use different articulators, but also of gestures that share articulators so that the resultant articulatory movement is a kind of complex vector with recoverable components of movement for each gesture. For example, when a talker coarticulates a bilabial stop, /b/ for example, with a vowel, say /a/, the tongue backing gesture for /a/ will begin during the closing or closure phase of the stop. This gesture does not share any articulators with the closing gesture. However, the segments do share at least one articulator, the jaw. It contributes both to the closing gestures for /b/ and to the opening gesture for /a/. Even so, the two gestures coarticulate so that during closure of /b/, the jaw is more open in /ba/ than in /bi/ (unpublished data of Keating, Lindblom and Lubker, cited in Keating, in press). Accordingly, the lips must do correspondingly more of the work to achieve bilabial closure in the context of /a/ than of /i/. The jaw closing movement may be seen as the product of two influences on it, influences working to close the jaw for /b/ and others, initially weaker owing to the later serial position of /a/, working to open it. By hypothesis all of the phonetic information that the talker intends to convey is in these overlapped articulatory gestures. None is destroyed in a translation of planned phonological segments from mind to mouth.

Certainly, the serial ordering of phonetic segments is maintained in speaking despite the overlap of gestures for different segments. Central to the identification of a segment as a consonant is a gesture creating a constriction in the vocal tract; central to identification of a vowel is an opening gesture. The serial ordering of closing and opening gestures in a sequence of phonetic segments is strictly maintained. In the example of /ba/ above, even though the closing and opening gestures for the consonant and vowel overlap, bilabial closure is achieved before opening for /a/.[8] The gestures that begin early do so relative to the closing or opening gesture for the phonetic segment that they, in part, constitute. A listener can extract the serial ordering of segments by tracking the sequence of their closing and opening gestures.

It remains a question—as yet unexplored by production researchers—how talkers maintain and provide information for the cohesion of nonsynchronous gestures constituting a phonetic segment and the separation of overlapping gestures for different segments. Presumably, research on functional coupling among articulators briefly alluded to earlier will help to determine how talkers achieve cohesion and separation of gestures. (That is, coupling may be stronger among gestures for a segment than among gestures for different segments.) Clearly, they do achieve cohesion, as their spontaneous speech errors show. Talkers are vastly more likely to make a phonetic error in which the identity of originally intended phonetic segments is preserved (e.g., "pade mossible" for "made possible"; Shattuck-Hufnagel, & Klatt, 1979), but segments are exchanged, for example, than errors in which features of segments are exchanged (e.g., "glear plue sky" for "clear blue sky" in Fromkin, 1973) altering the identity of planned segments (Shattuck-Hufnagel & Klatt, 1979). Research on the acoustic consequences of functional couplings may show how talkers *signal* cohesion and separation to listeners. That they do so is evident from perceptual studies such as those reviewed below under "Listener-talker attunements."

There is another kind of articulatory layering in speech production that is not generally considered, along with coarticulation, as distortion or smearing. However, under a view of coarticulation as layering of gestures for different segments, it is strictly analogous. It is a layering of "suprasegmental" information with other suprasegmental and phonetic segmental information in a speech utterance. Talkers signal stress, intonation and speech rate, for example, using the same laryngeal and supralaryngeal articulatory structures that they use to produce phonetic segments, and so all of these sources of linguistic information are superimposed in patternings of the same acoustic variables that convey phonetic-segmental information. Talkers provide intonational meaning (e.g., they signal that their utterance is a statement, a question, a list, etc.) by varying the rate at which their vocal folds open and close (and so by varying the fundamental frequency (f_0) of the voice) in systematic ways over the course of an utterance. However, other information is carried by f_0 as well, including information for vowel height, consonant voicing, stress and early-to-late position in a phrase (by means of "declination"—a tendency for f_0 to decrease exponentially early-to-late in a phrase).

As in coarticulation of phonetic gestures, the various influences on rate of vocal-fold opening and closing are nonsynchronous and overlapping. Effects of vowel height and stress are local, effects of declination are more global and effects of intonation include both global and local patterns.

In short, a hallmark of speech is temporal and spatial layering--a highly efficient way to provide information as long as the layering is done in a way that allows reliable and accurate unravelling by listeners. In the next two sections of the chapter, I review evidence showing that listeners recover articulated phonetic segments in perception. Presumably, their doing so is exactly what allows them to handle acoustic consequences of articulatory layering as reviewed in the following section.

Perceiving gestures: Six kinds of evidence

1) Different signals, same place of articulation, same percept.

This example has already been described. It is the finding that listeners identify the syllables in Figure 1 as /di/ and /du/, with the same initial consonant, even though the acoustic signals providing information for place of articulation are different. The articulatory gestures producing the /d/s invariantly include a closing and opening gesture of the tongue tip against the alveolar ridge of the palate.

2)The same acoustic signal, different places of articulation and different percepts.

A complementary example also derives from early research by investigators at Haskins Laboratories (Liberman, Delattre, & Cooper, 1952). The same synthetic stop burst (centered at 1440 Hz) is identified as /k/ before steady-state formants for /a/, but as /p/ before /i/. The reason for the change in place identification has to do with information provided by the burst in relation to the resonances for the following vowel. When the vowel is /a/, the second formant for the vowel is the "front cavity" resonance--that is, it is the resonance of the vocal-tract cavity in front of the tongue-body hump for /a/ (e.g., Kuhn, 1975; Stevens & House, 1956). When the burst is near that resonance, as it is at 1440 Hz, a consonantal place of articulation is specified that is near the place of the vocalic constriction. Therefore, the consonant must be a velar consonant, here /k/, which also uses the tongue-body to make its constriction. When the vowel is /i/, F3 is the front-cavity resonance, and the burst is far from that resonance, so it cannot be the burst from a velar consonant. In productions of bilabial stops, there is, initially, no front cavity because the foremost constriction is at the lips. On release of the closure the front cavity is the cavity in front of the constriction for the vowel. As the mouth opens, there is a rapid rise in the front cavity resonance to F3 if the vowel is a front vowel such as /i/. Listeners hear the burst as the starting point for that frequency rise and identify the consonant as a bilabial.

3) Recovery of cavity-affiliations of F2 and F3

In related research, Kuhn (1979) has found that listeners recover the cavity affiliations of vocal-tract resonances. In speech syllables, the second and third formants each may be associated with either the cavity behind a constriction in the vocal tract or the cavity in front of the constriction. (In general, whichever cavity is larger will be associated with the lower-frequency resonance, and that, by convention, is labeled F2.) Knowing where a constriction is in the vocal tract requires knowing more than just the relative sizes of the cavities, it requires knowing which cavity (the larger or the smaller) is in front of the constriction. If listeners recover place information by recovering constriction gestures, then, it seems, they must recover the cavity affiliations of the formants. In contrast, if they recover place information from the signal without using the signal as a source of information about articulation, then the varying cavity affiliations of F2 and F3 will not be relevant to recovery of phonetic information. Research by Kuhn (1979) and by Dorman, Studdert-Kennedy, and Raphael (1977) suggests, however, that the formant that constitutes the front cavity resonance, whether it is F2 or F3, provides the most salient information for place to listeners.

4) Functional equivalence of different acoustic "cues" specifying the same phonetic gesture

Some articulatory gestures have multiple acoustic consequences. Production of the closing/opening gesture for a stop consonant may yield (in temporal order): a) a closure interval during the consonantal constriction that is silent, or if the stop is voiced, that includes low amplitude voicing; b) a burst at release of the constriction; c) a period of frication at the constriction location; d) a period of aspiration produced at the glottis; and e) formant transitions into the next phonetic segment (see e.g., Fant, 1973). Because these are joint products of a closing/opening gesture, they are not, in fact, independent. However, they can be manipulated experimentally as if they were.

When this is done, one very striking finding is that acoustically quite distinct cues—say a silent interval and transitions into a following segment appropriate for a bilabial stop—may be functionally equivalent. In the proper context, either one may specify a stop. Even more striking is evidence that, in context, the cues may be nearly indiscriminable. Fitch, Halwes, Erickson, and Liberman (1980) found that a pair of syllables differing in two cues—in duration of a silent interval between /s/ frication and the remainder of a syllable, "-lit" and in transitions into "-lit"—will either be more discriminable than a pair of syllables differing in just one of these ways or *less* discriminable, depending on how the cues are combined. If one syllable has a relatively long silent interval and transitions for a bilabial stop, and the other has a shorter silent interval and no transitions, then the first syllable will sound like "split" and the second like "slit" and they will be highly discriminable (and more discriminable than a pair of syllables differing in just one of these ways). In contrast, if one syllable has a long closure and no transitions while the other has a shorter closure and transitions, both may sound like "split" and be nearly indiscriminable (and less discriminable than a pair of syllables differing in just one of these ways). Thus, syllables differing acoustically in two ways may be less discriminable than syllables differing acoustically in one of these ways, just when the two cues are joint consequences of production of the constriction and release gesture for the same phonetic segment /p/.

5) Cross modal perception of gestures

The preceding example was one in which two acoustic cues, silence and formant transitions, that are quite distinct out of context, may be perceptually indiscriminable in contexts in which each specifies production of the same consonantal gesture. The final example is similar, but in this case, the distinct "cues" are acoustic and visual. A videotaped face mouthing /ga/ synchronized with an acoustic /ba/ may be *heard* as saying /da/ (MacDonald & McGurk, 1978). In isolation, the visual and acoustic events are quite distinctive of course. Together, they are integrated to specify a common event in which the observer has great difficulty, and generally cannot, tell what part of the percept was contributed by the visual modality and what part by the auditory modality (Liberman, 1982 and research by Massaro and his colleagues summarized in Massaro, 1987).

Research by Green and Miller (1985) shows that integration of information for an utterance across modalities occurs also when the information is, in part, suprasegmental. Green and Miller (1985) report that the rate, slow or fast, at which a bilabial stop-vowel syllable is mouthed on a videotape will affect the boundary along an acoustic voice-onset time[9] continuum between /p/ and /b/ in /bi/-/pi/ syllables spoken at an intermediate rate. In particular, listeners heard more /pi/ syllables when the visually-specified rate was fast than when it was slow. In this experiment, without the acoustic signal, viewers could not identify the mouthed syllable as /bi/ or /pi/; without the video display, they heard the syllables at the acoustically-specified intermediate rate. Information for segment identification clearly was integrated across modalities. The visual modality provided information for rate needed to evaluate each voice-onset time value as characteristic of /p/ or /b/; the acoustic signal provided the voice-onset time values.

These findings show that listeners integrate information for articulatory gestures not only across acoustic cues, but also across relevant modalities. Although the findings have been given a different interpretation than I give them here (Massaro, 1987), in my view, they provide striking evidence that listeners recover distal articulatory events in speech perception from either informational medium, optical or acoustic, that may have been structured by those events.

6) Shadowing

Listeners asked to repeat ("shadow") a spoken syllable as fast as possible in a choice response time paradigm exhibit latencies nearly as short as they show in a simple response time procedure and much shorter than they show in a choice reaction time procedure with a button-press response (e.g., Porter, 1978; Porter & Lubker, 1980). One interpretation of the extraordinarily short times to repeat a spoken syllable is that listeners extract information specifying how the talker produced the syllable. The only response that involves essentially no choice at all, and so can be very fast, is a repetition.

Listener-talker attunements

The foregoing sources of evidence suggest very strongly that listeners recover talkers' linguistically-structured vocal-tract actions. Even if listeners do that, however, it appears that they should still be confronted with a fairly severe problem of pattern recognition if they are to recover phonetic segments from the vocal-tract activity. Because of articulatory layering, a given vocal-tract action—the rate of vocal fold opening and closing, for example, or the opening or closing of the jaw—reflects the convergence of multiple influences. To recover the phonetic segments, the listener must somehow perceive these various influences on a common articulatory behavior as separate. Too, gestures are not phonetic segments; a given phonetic segment is realized jointly by several gestures that, as described earlier, are temporally staggered. To recover the phonetic segments, listeners have to perceive the integrity of gestures that together specify a consonant or vowel. In short, where appropriate, they have to perceive the separation of multiple influences on a common articulatory movement and/or a common acoustic variable and the cohesion of temporally staggered gestures for a common phonetic segment. The literature on perception suggests that this is exactly what they do. As listeners, we discover more about what talkers are doing that we have succeeded in discovering as speech researchers. Following are some examples.

Unravelling multiple influences on f_0

F_0 reflects the rate at which the vocal folds open and close for voiced consonants and vowels. Physiologically it may be increased by increasing air pressure below the vocal folds ("subglottal pressure" or Ps) or by tensing the vocal folds. The vocal folds may be tensed directly by contracting appropriate muscles in the larynx (the cricothyroid or vocalis muscles) or indirectly by forces that pull or rotate the laryngeal cartilages supporting the vocal folds. As noted earlier, linguistically, f_0 is used to realize the intonation contour of a sentence and it is a marker of stress. In addition, f_0 declines exponentially over the course of a major syntactic unit such as a sentence, and it shows systematic perturbations due to vowel height (with f_0 increasing with vowel height) and to the voicing of a neighboring obstruent. How can the listener know whether a given high value of f_0 signals an intonation peak, the start of a declination contour, a stressed syllable, a high vowel or a preceding voiceless obstruent? They do know, as the examples below illustrate.

Declination and intonation. Pierrehumbert (1979) found that listeners identify the later of two intonational peaks in a sentence as higher than the earlier peak if the two peaks have the same value. The peaks sound equal only if the later one is lower. Listeners behave as if they are parsing the f_0 contour into its regular decrease due to declination and the f_0 excursions relative to it that constitute the intonation contour. Equivalent f_0 excursions will reach lower peak values later in a sentence than earlier ones because the f_0 baseline is lower due to declination.

The listener's parsing of the contour appears to mirror the relative independence of declination from the intonation contour in production. Declination largely reflects the decline in P_S over a sentence due the the deflation of the lungs (Gelfer, 1987; Gelfer, Harris, & Baer, 1987), while intonational peaks are generally ascribed to increases in vocal-fold tension (e.g., Breckenridge, 1977; Ohala, 1978).

What tells the listener to parse the contour in this way? Two kind of information are available in the signal. One is that, whereas intonational peaks are relatively local changes in f_0, declination is a more global change in the average f_0 value. The two sources of influence on a peak have different time scales. A second source of information that may support the listeners' parsing of the f_0 contour is that, in natural speech, declination of f_0 is accompanied by declination of the amplitude contour (Breckenridge, 1977; Maeda, 1976). Presumably, both variables decline because P_S declines, and both variables reflect P_S. Listeners can use the joint fall in f_0 and amplitude to pinpoint P_S as the articulatory source of the fall, and they can use that information to pull out the contribution of P_S to f_0.

Intonational peaks and intrinsic f_0. Listeners parse segmental influences on f_0 from the intonation contour as well. Contraction of the posterior fibers of the genioglossus muscle to produce forward movement of the tongue root for high vowels pulls the hyoid bone of the larynx forward. In turn, the hyoid rotates the thyroid cartilage stretching and tensing the vocal folds and causing f_0 to rise (Honda, 1981). Therefore, high vowels have higher fundamental frequencies than low vowels. Like declination, this effect on f_0 is superimposed on the intonational patterning, but, in contrast to declination, its effects are local to a vowel.

Silverman (1985; 1987) has shown that listeners hear equal f_0 peaks as different if one occurs on a low vowel and one on a high vowel. He found that the magnitude of the "adjustment" to the perceived peak is about the same as the effects that the low and high vowels in fact do have on f_0.

How do listeners know what part of the f_0 pattern is due to the pull on the hyoid and thyroid cartilages? One possibility is that they use spectral information for the vowel to recover tongue movement, which implicates pull on the larynx. (Although it is possible to achieve similar positions of the tongue relative to the palate with various contributions of tongue and jaw (as in bite-block speech; see e.g., Lindblom, Lubker, & Gay, 1979) and hence with consequences for laryngeal pull, there are not many options (Ladefoged, 1980; Ladefoged, Harshman, Goldstein, & Rice, 1978) and among the possible ones, most are unlikely, because they are effortful.)

Intonational peaks and obstruent voicing. A vowel following a voiced or voiceless obstruent has a high falling tone on it (Ohde, 1984), apparently an unintended automatic consequence of articulatory maneuvers for the consonant (e.g., Ohala, 1978). The tone following a voiceless obstruent begins higher than that following a voiced obstruent. Listeners use the tonal contours as information for voicing (Haggard, Ambler, & Callow, 1970), and they factor it from a superimposed intonation contour (Silverman, 1986).

In Silverman's research, an intonation contour on a vowel preceding a stop consonant either rose or fell. Following the silent closure for the consonant, in one condition, Silverman introduced a brief level contour on the next vowel followed by a continuation of the rise or fall. (See Figure 2.) The level contour was identical in f_0 in utterances with overall rising or falling intonations. However, in the context of a rising contour it was higher in frequency than f_0 would have been had the rise continued along the path it took into the stop closure. In the falling contour it was lower than the frequency of the contour interpolated through the closure. Accordingly, for the listeners who factor the contour from perturbations due to the consonant, the level interval in the rising contour signaled a f_0 perturbation causing f_0 to rise, and hence signaled a voiceless stop. The perturbation in the falling contour signaled a voiced stop. Listeners' judgments fit the predicted pattern.[10]

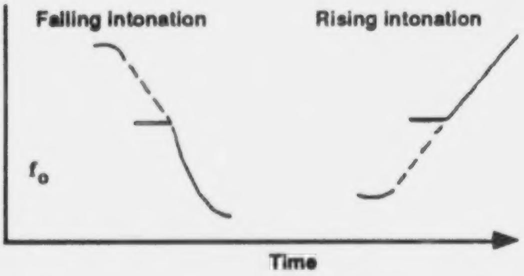


Figure 2. Falling and rising intonation contous with level f_0 perturbations used by Silverman.

Unravelling multiple influences on duration

Durations of articulatory gestures and measured acoustic durations of phonetic segments also are determined by multiple linguistic and paralinguistic variables. Segments have "intrinsic durations" (Lehiste, 1970)—at the same rate of speech, some are longer than others. Durations generally shrink with speech rate, and, like any other system braking to a stop, articulated segments lengthen at boundaries (sentence boundaries, for example). In addition, they lengthen when stressed, and stressed vowels shorten in the context of following consonants or unstressed syllables (a phenomenon called "compensatory shortening").

Listeners handle these converging influences on duration as well as they handle effects on f_0.

Rate and intrinsic duration. In synthetic CV syllables, /b/ and /w/ may be distinguished by the durations of formant transitions into the following vowel; those for /w/ are longer than those for /b/. Transitions for /w/ also shorten with increasing speech rate, and Miller and Liberman (1979) have shown that a boundary between /b/ and /w/ on a transition-duration continuum shifts to a shorter duration if vowel durations are shortened, specifying a faster speech rate. Just as listeners factor out a global effect of declination from local intonational peaks, so in perception of segments, they extract relatively global information for rate (here, syllable-wide) from local durational information for segment identification.

Interestingly, the effect of rate is due only to consequences of rate changes physically present in the signal, not to consequences of subjectively experienced changes in rate if those subjective experiences are not backed up by gesturally-produced rate changes. Miller, Aibel, and Green (1984) embedded stimuli from a /ba/-/wa/ continuum in the context of slow- or fast-rate syllables thereby inducing a contrast effect in judged rate. (That is, stimuli in the context of slow-rate syllables were judged as faster than the same syllables in the context of fast-rate syllables.) However, this manipulation had no effect on the /b/-/w/ boundary. That visually-specified rate *does* shift the boundary (Green & Miller, 1985, described earlier), whereas subjective rate does not, indicates in another way that visual and acoustic information for a speech event can serve equivalently as effective information for listeners. Presumably the two kinds of information are equivalent because of their common gestural source; both specify the rate of articulatory gestures in speech.

Compensatory shortening, final lengthening and intrinsic duration. The Dutch language has phonologically distinct vowels that differ only (or largely) in duration. For example, /a/ and /aː/ are distinct vowels in the language, distinguished acoustically and articulatorily by the longer duration of the latter segment. Nooteboom (1975; cited in Nooteboom, Brokx, & De Rooij, 1978) found that a boundary in a continuum from /a/ to /aː/ is shifted to shorter durations in contexts that induce compensatory shortening. Accordingly, a stressed vowel identified as /a/ in isolation may be identified as /aː/ followed by /s/ or /t/. Likewise, a stressed vowel identified as /a/ followed by /t/ is progressively more likely to be identified as /aː/ before one, two or three unstressed syllables.

Likewise, a word identified as /taːk/ in the middle of a sentence may be identified as /tak/ sentence-finally where, as a talker slows to a stop, segments are subject to final lengthening.

Summary

In all of the examples presented so far, listeners pull apart the multiple sources of influence on a common acoustic variable in a way that recovers their separate linguistic and paralinguistic sources. A hypothesis that they do so because they have stored vast amounts of information on the acoustic consequences of gestural layering leaves unanswered how they could ever discover that information before they had it. Too, in the one case that has been looked at in relatively inexperienced perceivers (Eimas & Miller, 1980; Miller & Eimas, 1983), three to four month old infants parsed rate information from information for segment identification in the same way that adults do; this suggests that the knowledge is not acquired. It is either innate knowledge or it is in the signal (or both). If we hypothesize that it is innate and not in the signal, it seems to me that we have the same puzzle left unanswered by the learning hypothesis. How did the information get to be innate if it were not already there in the world to be discovered and used?

We know from studies of visual perception that observers discover what components of an event belong together and what components are separate in part by noticing how the components undergo change. Point lights affixed to the joints of a moving person filmed in the dark (e.g., Johansson, 1973; 1974) readily specify both the form (a human figure) and the nature of its movement (e.g., walking). Viewers can extract the form from motion by detecting how the parts move. For a walker moving left to right across an observer's field of view, all of the point lights for the figure have a vector of left-to-right motion. In addition, points on a leg, for example, have a back and forth motion relative to the common motion. That the points on the leg move together specifies their connectedness; that they share a motion vector with the other point lights too specifies their connectedness to those points as well, but their motion relative to those other points specifies a degree of separation. Johansson (1973; 1974) has shown that viewers use common motion vectors as evidence for cohesion, relative motion as information for a degree of separation, and no motion in common at all as evidence for independence.

I suggest that speech perception works similarly. Common motions for the components of an utterance may be provided by global events that span long stretches of an utterance—among others, the declination contour and information for rate. More local excursions relative to those global events--for example, production of an intonational peak or producing a /w/ are perceived as "relative motions"—relative that is to the global events outlined by a common speech rate or declination line.

There is other information for cohesion as well—in particular, information provided by a cluster of acoustic events that jointly and redundantly specify a particular articulatory event. The declination line can be identified as such because it falls, but also because it falls in concert with the amplitude contour. This pinpoints P_s as its articulatory source and separates it from any other influences on f_0 that do not involve changes in P_s. By the same token, local increases in f_0 that are accompanied by information that the tongue is up near the palate—a low F1 for example—can be identified thereby as increase in f_0 due to stretching of the vocal folds by forces exerted on the hyoid bone and thyroid cartilage (Honda, 1981).

For its part, final lengthening will not be confused with durational length that signals intrinsic duration or with stress or with a general slowing in speech rate because it is accompanied by other consequences of articulatory stopping: sharp falls in f_0 and in amplitude.

Perceiving cohesion and separation of gestures for coarticulating phonetic segments

I have suggested that coarticulatory overlap is a gestural layering of the same general sort as the layering of segmental and suprasegmental gestures in speech. If so, then listeners should be able to perceive the separateness of coarticulating segments and the cohesion of the temporally staggered parts of a segment in the same way that they perceive the separation of declination from intonational peaks, for example, and the cohesion of the temporally spread declination line. They should fail to integrate so-called "anticipatory" and "carryover" acoustic effects of one segment with the acoustic effects of gestures for another. In addition, the anticipatory and carryover effects should be perceived, but as integral parts of the coarticulating segment itself.

Two series of experiments show exactly this pattern. In one (Fowler, 1984), stop bursts from natural tokens of /gi/ and /gu/ were extracted from their original context. They were spliced onto the vocalic portions of new syllables. In the "spliced" condition, a stop burst from one token of /gi/ (/gu/) was spliced onto a vocalic portion of another /gi/ (/gu/) syllable. In the "cross-spliced" condition, the stop burst from a token of /gi/ was spliced on the the vocalic portion of /gu/ and vice versa. Velar stop bursts include information for a following vowel. Accordingly, in the cross-spliced condition the burst includes misleading information for the vowel to come. Similarly, the vowel misleads as to the acoustic information in the consonant for which it is responsible, due to anticipatory coarticulation.

In a second series of experiments (Fowler & Smith, 1986), the stimuli were trisyllables (/ibəbi/, /ibəba/, /abəbi/ and /abəba/). The medial /bə/ syllable was spliced and cross-spliced within and across flanking vocalic contexts. In cross-spliced conditions, the unstressed vowel contains misleading information for the flanking vowels and they mislead as to the acoustic information in the schwa[11] for which they are responsible due to coarticulatory overlap.

Two experiments were performed using each stimulus set, one designed to ask whether listeners perceive coarticulatory influences of one phonetic segment on another as separate from the coarticulated segment. The other asked whether they integrate that information with other information for the coarticulating segment. To test for separation, we used a 4IAX discrimination procedure. Subjects heard four stimuli on each trial, presented in two pairs. They were asked to judge in which pair the members sounded more alike one to the other. A sample trial from the second stimulus set is presented below. Subscripts identify the original context for the unstressed syllable:

i_ib $ə_i$bi—a_ib $ə_i$ba——i_ib $ə_i$bi—a_ab $ə_a$ba

Subjects judged whether the first two trisyllables sounded more or less alike than the second two. The task is difficult because the trisyllables are quite dissimilar. However, the task can be done if listeners attend to the medial syllables. In the first pair, the medial syllables are identical, but one is spliced and the other is cross-spliced into a context of vowels different from the one in which it was produced originally. In the second pair the medial syllables are different, but each is spliced into a context like the one in which it was originally produced. If listeners perceive the gestures for /a/ and /i/ that begin and end in the medial syllable as separate from the /bə/ syllable, then the medial syllables in the last pair should sound identical, while those in the first pair will have different coarticulatory information removed from them and will sound different.

In both experiments, the one examining vowel-to-consonant coarticulatory influences and the other examining vowel-to-vowel effects, the results were the same. Listeners' discrimination judgments suggested that they had pulled out the coarticulatory influence of a neighboring segment from the segments being judged. Stop consonants (and unstressed vowels in trisyllables) in the context of different stressed vowels were judged more alike if the signals specifying them were acoustically different than if they were identical—presumably because differences between members of a pair were eliminated and created respectively by extraction of coarticulatory influences in perception.

A second experiment performed with each stimulus set was designed to ask what happened to the coarticulatory information factored from the /g/ burst or /bə/ syllable. We hypothesized that listeners would perceive coarticulatory information as part of the coarticulating segment. If so, accurate anticipatory coarticulatory information in the /g/ burst or in the unstressed syllable should facilitate identification of the coarticulating segment; misleading coarticulatory information should slow its identification (see also Whalen, 1981; 1984). In the experiment, listeners participated in a choice-reaction time task. In the version using trisyllables, listeners heard one of /ibəbi/, /ibəba/, /abəbi/ or /abəba/ on each trial with the medial syllable either spliced or cross-spliced. Their task was to identify the final vowel as /i/ or /a/ as quickly as possible. In the version using stop-vowel syllables, stimuli were spliced and cross-spliced /gi/ and /gu/; listeners identified the final vowel as /i/ or /u/ as quickly as possible. In both experiments, response times were slowed in the cross-spliced conditions as compared to the spliced condition presumably because cross-spliced stimuli provided misleading coarticulatory information for the segment being judged.

Across the pair of experiments, results suggest that listeners recover coarticulated segments as talkers produce them. Talkers produce a gesture of the tongue body for a stressed vowel that begins during the /g/ in isolated /gi/ or /gu/ and during the unstressed vowel in the trisyllables. The gesture extends to, and continues during, jaw lowering for the vowel. Listeners behave as if they detect the cohesion of the extended tongue movement for the vowel and, thereby, detect the separation of that gesture from gestures for the consonant (or unstressed vowel) with which it cooccurs in part.

A final, quite remarkable, experiment by Mann (1986) suggests a similar conclusion. As described earlier, Miller et al. (1984) showed that effects of rate on identification of a segment depends on information for rate physically present in the signal (acoustic or optical) not necessarily on phenomenal rate as judged by listeners. Mann reports a similar finding as regards perception of coarticulated speech. In earlier work, Mann (1980) had found that the boundary along a /da/-/ga/ continuum is shifted by a preceding /al/ syllable as compared to a preceding /ar/. /al/ led to more "g" identifications. This is expected if listeners pull out consonant-to-consonant coarticulatory effects in the way that experiments by Fowler (1984) and Fowler and Smith (1986) suggest they pull out vowel-to-consonant and vowel-to-vowel effects. /l/ has a more forward point of articulation than /r/ and it pulls the constriction place for /g/ forward. Listeners detect that fronting effect of /l/ on continuum members and, hence, they identify more fronted continuum members as "g" in the context of /l/ than in isolation or in the context of /r/. Interestingly, Mann (1986) found that the context effect was indistinguishable among English and Japanese listeners--even among Japanese listeners who were nearly at chance in identifying the syllable-final consonants in the disyllables as /l/ or /r/.[12] Mann concludes that the listeners' factoring of the coarticulatory effect of /l/ on the continuum members must be supported by information for coarticulatory gestures. Information for these gestures in the acoustic speech signal is as available to Japanese as to English listeners. Indeed, the study suggests very strongly that information for the coarticulating gestures must be in the signal, because, in their identification judgments, many of the Japanese listeners demonstrated that they did not have the stored knowledge needed to name the coarticulating segments reliably as "l" or "r." Information supporting the listeners' naming of the coarticulating segment must be mediated by the different phonological categories of the the native languages.

Attunement of listeners to phonetic gestures: a final argument from the literature on source separation.

I have attempted to show, in two domains, perception of superimposed suprasegmental and segmental information and perception of coarticulated segmental information that listeners recover patterns of cohesion and separation that reflect what talkers do gesturally. When a decrease in f_0, generally accompanied by a decrease in amplitude, spans a sentence, listeners detect its coherence over time and ascribe it to a different source than local f_0 movements due to other factors. And that exactly reflects the distinct articulatory sources of declination and other sources of systematic variation in f_0. When talkers begin a tongue raising and fronting gesture for /i/ during a preceding consonant or unstressed vowel and, of course, continue it during jaw opening for the vowel, listeners detect the coherence of that gesture and its cohesion with other gestures for the vowel, and they detect the separation of that gesture from the local excursions of the tongue for an unstressed vowel or of the same or some other articulator for a consonant that cooccur with part of its trajectory.

To many researchers, this implies that perceivers must have a large storehouse of knowledge, learned or innate about the acoustic consequences of superimposed articulations. I have argued that, whether or not they do, the information supporting acquisition and use of that knowledge must be in the stimulation to perceivers.

The same conclusion can be argued in a different way by comparing perceivers' handling of superimposed speech and audible nonspeech events. This kind of acoustic mixing cannot be anticipated by evolution or learning, because, as the two series of experiments cited below show, novel audible nonspeech events may overlap with speech and, when they do, listeners unravel their effects, apparently in just the way that they unravel multiple gestural influences on common articulatory or acoustic variables.

Darwin and Sutherland (1984) created an acoustic continuum from /I/ to /e/ by varying F1 in steps from 375 Hz (appropriate for /I/) to 500 Hz (characteristic of /e/). They obtained identification judgments of continuum members under several conditions. In one, the continuum members were presented in isolation; in another, a 500 Hz tone was presented, cotemporally with each vowel; in others, a 500 Hz tone began some time before the vowel and continued throughout the vowel. If the 500 Hz tone is perceptually integrated with the vowel, it will increase /e/ judgments. If it is perceived as a separate nonspeech event, it will not affect vowel identifications. Darwin and Sutherland found an increase in /e/ judgments only if the tone was wholly coincident with the vowel so that there was little or no information for the tone's separation from the vowel. If it led the vowel, even by as little as 32 ms, the tone did not contribute to vowel identifications at all. The effect of the leading tone could be partially reinstated by "capturing" the leading edge of the tone with a 1000 Hz tone simultaneous with it.

Here is an example in which separation of different distal sources and integration of information for a common distal source each appears to proceed just as it does when the sources both reside in the vocal tract of a talker. But in this case, one source is in the vocal tract and the other is not.

"Duplex perception" may provide an example of the same sort, although it is not given that interpretation by its investigators (e.g., Isenberg & Liberman, 1978; Whalen & Liberman, 1987). In the usual procedure for obtaining duplex perception, all of a CV syllable (say, /la/ or /ra/) is presented to one ear except for the F3 transition, which is presented to the other ear. (The transition is the only information in these syllables distinguishing /la/ from /ra/.) The major part of the syllable is called the "base" and the transition, the "chirp." Heard in isolation, the base is identified as /ra/ and the chirps sound like rapid changes in frequency. When the base and /l/ chirp are presented dichotically, listeners hear /la/ in the ear that receives the base and, simultaneously, they hear the /l/ chirp in the other ear. Hence, perception is "duplex." This outcome has been described as one of hearing the same signal (the chirp) simultaneously in two ways (e.g., Liberman & Mattingly, 1985; Mann, 1986; Whalen & Liberman, 1987). In turn, this is seen as evidence by Liberman and Mattingly that speech perception involves a separate neurological mechanism from perception of nonspeech and that, in some cases, both systems can be tricked into rendering a percept including the chirp. However, recent findings by Whalen and Liberman (1987) suggest to me that this interpretation is unwarranted and, indeed, that duplex perception provides another example of perceptual "vector analysis" involving superimposed speech and nonspeech events.

Whalen and Liberman showed that dichotic presentation is not required for duplex perception. They presented sinewave versions of a base and chirp diotically, but at different intensity levels. In particular, the chirp's intensity was varied from a level at which listeners could not hear it at all to a level at which they experienced duplex perception. There was an intermediate range of intensities at which subjects integrated the chirp with the syllable, but did not hear the chirp.

This finding suggests that duplex perception is similar to that reported by Darwin and Sutherland (1984), now with intensity playing a role somewhat similar to that which temporal asynchrony played in Darwin and Sutherland's research. In natural speech, the chirp will bear some intensity relation to F3. Possibly, on average it will be low in intensity initially, with intensity gradually increasing as the mouth opens. In any case, there will be some range of intensities of the chirp that will fit the intensity of F3 of the base and a range of intensities that are too high. When the intensity is too high, listeners hear the residual, which does not fit the syllable, as a chirp.

I suspect that, whether superimposed audible events are both (or all) nonspeech (for example, a telephone ringing and a door closing), one is speech and one nonspeech (as in Darwin and Sutherland's study) or both are speech (as in the research summarized under "Listener-talker attunements"), the perceptual process is the same. Listeners use information in the signal for separation to detect the separation of overlapping events and information for cohesion to discover what belongs together. There may be nothing special about this process applied to overlapping events taking place in the vocal tract.

Listeners as sources of sound change: Failures of attunement?

Languages undergo orderly and continual changes in pronunciation of words. Ohala (e.g., 1974; 1981) has suggested that systematic mishearings by listeners may explain some of these changes. In particular, he suggests that listeners may sometimes misparse an acoustic speech signal, failing to ascribe coarticulatory influences to an influencing segment, but instead integrating them with the influenced segment (see footnote 6). When this happens systematically in a language community (among language learners, presumably), a sound change may eventuate.

Ohala (1974) suggests as one example of this the development of tones in certain tone languages. In these languages, (e.g., Punjabi, Chinese and many South East Asian languages) tones developed when a voicing distinction was lost. Instead of the voicing difference, words previously having a voiceless consonant now had a high tone on the following vowel and words having had a voiced consonant now had a low tone on the next vowel.

One account of the development of tones in these languages (Hombert, Ohala & Ewan, 1979; Ohala, 1974) ascribes it to a misparsing of the signal. As the voicing distinction weakened, so did information in the acoustic signal telling a perceiver that the tonal pattern on the vowel had been perturbed by the preceding consonant. Accordingly (in contrast to listeners in Silverman's study (1986) for whom the voicing information was clearly audible), listeners may have begun to hear the tonal pattern as part of the vowel. Language learners perceiving the words that way would be inclined to produce a tonal pattern on the vowel intentionally, exaggerating it in doing so. Eventually, the tonal pattern provides the distinctive information in these words; the voicing difference is further weakened and ultimately lost.

In this way, a pattern appears in the phonology of the language that resembles an acoustic byproduct of an articulatory gesture. But it is intentionally produced and it now "belongs" to the segment with which it cooccurs in time (the vowel, in the example), not to the segment (the voiced or voiceless consonant) that caused its appearance when it was a gestural by-product.

In fact there are many examples of this sort that may reflect consequences of inexperienced listeners attempting to extract information for articulation from the signal. A few examples follow that relate to some of the examples of "attunements" described above under "Listener-talker attunements."

1. f_0 and vowel height. f_0 is correlated with the height of a vowel, as noted earlier, probably because contraction of the posterior fibers of the genioglossus muscle to raise the tongue also pulls on the hyoid bone, rotating the thyroid cartilage and stretching the vocal folds (Honda, 1981). There is at least one example of a phonological rule that may have its origins in this correlation. It is described by Mohr (1971) and Wang (1972).

In the Foochow dialect of Chinese, morphophonological rules raise tones on a vowel in the context of various following tones. Other rules raise the vowel on which a tone has been raised. Both Mohr and Wang suggest that the historical order of events was the reverse of this. That is, vowels were raised in certain contexts and then, eventually, their tones were raised as well. If they are correct, here is an example in which an articulatorily motivated correlation between vowel height and f_0 is elevated into the phonology of the language in stylized form (that is, as one of the tones of the language), perhaps via listener misparsing of the f_0 contour.

2. Compensatory shortening and compensatory lengthening. "Compensatory shortening," as described earlier, is a measured acoustic shortening of a vowel by neighboring consonants or of a stressed vowel by neighboring unstressed syllables. In the languages that show compensatory shortening, it tends to be asymmetrical. Vowels are shortened more by following than by preceding consonants (Swedish: Lindblom & Rapp, 1973; English: Fowler, 1983) and stressed vowels are shortened more by following than by preceding unstressed syllables (Swedish: Lindblom & Rapp, 1973; Dutch: Nooteboom, 1973; English: Fowler, 1981).

A common sound change in some language families, including Germanic languages, is called "compensatory lengthening" (e.g., Grundt, 1976; Hayes, 1987; Ingria, 1980). In compensatory lengthening a short segment, often a vowel, becomes phonologically long in the neighborhood of a lost consonant or unstressed vowel. (So for example, in Latin, /fideslia/ became /fide:lia/ and in Middle English, /tal / became ta:l/ [now English "tale"].) Hayes (1989) points out that the lengthening effect is strictly asymmetrical. Segments lengthen due to weakening or loss of following but not preceding consonants and schwas.

The similarity between compensatory shortening and compensatory lengthening cannot be accidental, but they are not the same phenomenon. In compensatory lengthening, a phonologically short vowel becomes a different vowel in the language--one that is phonologically long and that tends to differ somewhat in quality from the short vowel too (Grundt, 1976). In compensatory shortening, the vowel does not change identity, only acoustic duration.

Possibly the phenomena are connected in the same way that Ohala suggests tones in Punjabi are connected to f_0 perturbations on vowels due to obstruent voicing. As a consonant or schwa is reduced and lost in a language showing compensatory *shortening*, a neighboring vowel will lengthen acoustically. Normally (as Nooteboom's experiments on Dutch, described earlier, show), listeners do not hear duration differences due to different numbers of following consonants or unstressed vowels as duration differences. Possibly, however, as the weakening segment reduces but still remains audible, listeners will misunderstand the reasons for the progressive lengthening of the preceding vowel and will hear it as a true duration increase, characteristic of a phonologically long vowel.

3. Vowel-to-consonant assimilations. My research on vowel-to-consonant coarticulation (Fowler, 1984) showed that listeners extract coarticulating vowel information from a consonant (in /gi/ and /gu/ syllables) and ascribe the information to the vowel. Pagliuca (1982) describes a number of sound changes in languages that reflect what he calls "coalescence" of gestures for a consonant with coarticulating vowel gestures.

Some examples of common kinds of changes involve velar consonants being pulled forward in the context of more front vowels or lip-rounded vowels. In all of the languages in which this change has occurred, the velar consonant was undergoing a "weakening" over time (so, for example, a stop such as /g/ or /k/ became a fricative with the same place of articulation (/ɣ/ or /x/). The weakening may have rendered the consonants more susceptible to coalescence as the vowel information in the consonant perhaps became relatively more salient.

In Finnish, /k/ first weakened to /ɣ/. Between high rounded vowels, /ɣ/ became /v/ (so a voiced velar fricative became a voiced labiodental fricative), presumably a failure of rounding for the vowel to be extracted from the consonant.

In a Northern Lappish dialect called Karasjok, /k/ weakened to /ɣ/. It remained /ɣ/ before /a/, but became a labial glide, /w/, before (rounded) /u/ and /o/ and a palatal glide, /j/, before front vowels. Both changes suggest a coalescence of an articulatory gesture for a vowel (the rounding gesture of /u/ and /o/ and the fronting of front vowels) with the gesture achieving consonantal place and manner of articulation.

In Middle English, the voiceless velar fricative /x/ formerly at the end of such words as "cough", "laugh" and "rough" (and still represented in the orthography by "gh") became /f/. At the time of the change, all of the vowels in the words that underwent this shift in place of articulation were diphthongs ending in /u/. Once again, the fronting of the consonant may reflect a failure to extract all of the rounding influence of the vowel from the consonant.

4. Vowel-to-vowel assimilations. Vowels exert coarticulatory influences on one another even across intervening consonants (e.g., Fowler, 1981b; Öhman, 1966). Even so, listeners factor coarticulatory effects from an influenced vowel and ascribe them to an influencing vowel (Fowler and Smith, 1986).

There are at least two phonological processes that suggest occasional failures to achieve this factoring. One is a diachronic process called "umlaut" in German. Vowels that had been back vowels in Old High German split into two forms in Middle High German and Modern German (e.g., Bynon, 1977). For example, the back vowel [o:] remained [o:] in some contexts but shifted to the front vowel [ö:] in Middle High German in the context of a front vowel in the following suffix.

A second, now synchronic phonological process, that resembles vowel-to-vowel coarticulation is "vowel harmony." Vowel harmony is a common phonological process in languages, occurring for example in Turkish, Hungarian, Finnish, Mongolian, Nez Perce and Somali (Clements & Ezer, 1982). In Turkish, suffix vowels assimilate in backness and rounding to a preceding stem vowel.

In both phonological processes, umlauting and vowel harmony, a vowel assimilates along some dimension or dimensions with a neighboring vowel even if consonants intervene between the vowels. Possibly, these processes originate in misparsings of coarticulatory overlap among vowels.

5. F_0 downdrift. The phenomenon of declination has parallels in the phonologies of languages too, although they are unlikely to have developed by means of misparsing. As characterized earlier, declination is an exponential fall in f_0 (with more declination earlier than later) over the course of an utterance or utterance part. If intonational peaks (the "topline") are examined separately from troughs (the "baseline"), a difference in rate of fall can be observed, with the topline declining faster than the baseline.

One phonological analogue to declination is "downdrift" in some African tone languages (e.g., Hyman, 1973). In downdrift, a high tone on a word in a sentence is lowered following a low tone. In some languages, low tones are also stepped down, but this is less common and, where it occurs, it is less salient than the downstepping of high tones according to Hyman. Anderson (1978) describes the downdrift of high tones exponential in form.

Downdrift is not declination. Whereas a sequence of words with high tones does not exhibit downdrift, it presumably is subject to declination as the talkers' lungs deflate. Downdrift is a phonological stylization of a declination-like process.

According to Pierrehumbert (1980; M. Liberman & Pierrehumbert, 1984), English has a downstepping intonation contour, used for example on utterances of lists. In intonational downstepping, intonational peaks (high tones) are downstepped exponentially following intonation lows. The magnitude of downstepping exceeds that of declination and, like downdrift, it is stylized. High tones are downstepped only after lows and the intonation contour reveals true steps, not the gradual decrease in f_0 characteristic of declination.

Possibly these analogues to declination arose in a failure of listeners (unlike the listeners in Pierrehumbert's (1979) research) to pull f_0 declination from intonational peaks. Alternatively, their development may reflect a misapprehension of the source of declination in lung deflation. That is, language learners may have identified declination as something talkers imposed on the signal on purpose.

In fact, neither of these accounts is satisfactory, however, because both downdrift in tone languages and downstepping in English intonation is restricted only to some contexts. If declination were misperceived, it seems likely that it would be misperceived everywhere. Whatever their origin, however, these phonological processes remain examples of elevation of a gestural regularity common to all or most languages in exaggerated and stylized form into the phonologies of just a few languages.

SUMMARY

I have suggested that in fundamental ways, perceiving speech is like seeing. In both modalities, we perceive events taking place in the world and we do so by extracting information about those events from media that have been structured by those events. In speech, the distal events take place in the vocal tract, but they are by nature linguistic events. The structured medium providing the most information about vocal gestures having linguistic significance is the acoustic signal. Both vocal events and canonical visible events are complex. Lots of things are happening in overlapping time frames and so movements themselves tend to be complex motion vectors. Motion of the same body part (e.g., an arm of a walker, the jaw of a talker) may include vectors of motion owing to multiple sources (the path of locomotion of the walker and the arm swinging relative to that, the closing gesture for a consonant and the opening gesture for a coarticulating vowel). And motions of different parts may include common vectors because they are working together as an organized system (the whole body is moving forward; the jaw and lips are jointly achieving bilabial closure).

Perceivers behave as if they are perceiving distal events, but more than that, that they are discovering how the event is partitioned into separate and cohesive parts. In speech perception this means that the listener recovers the cohesion of gestures that, over time, jointly achieve a phonetic segment and they perceive the separation of gestures for different segments or for suprasegmental information, that are layered in articulation and hence in the acoustic speech signal. Indeed, research suggests that listeners are remarkably attuned to the talker's linguistic gestural activity.

I doubt that this way of perceiving speech is special; indeed listeners' detection that certain articulatory gestures are separate may be no different, fundamentally, from their detection that two audible events in the environment, one speech and one nonspeech for example, are separate.

The process may go astray sometimes, perhaps most likely among language learners. The presence of certain phonological rules in languages that mimic, often in stylized and exaggerated form, acoustic consequences of systematic articulatory processes suggests occasional general failures of listener-talker attunements.

REFERENCES

Anderson, S. (1978). Tone features. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 133-175). New York: Academic Press.

Blumstein, S., Isaacs, E., & Mertus, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation. *Journal of the Acoustical Society of America, 72*, 43-50.

Blumstein, S., & Stevens, K. (1979). Acoustic invariance in speech production: Evidence from measurement of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America, 66*, 1001-1017.

Breckenridge, J. (1977). *Declination as a phonological process*. Murray Hill, NJ: Bell Laboratories Technological Memo.

Browman, C. (1980). Perceptual processing: Evidence from slips of the ear. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*. New York: Academic Press, 213-230.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. In C. Ewan & J. Anderson (Eds.), *Phonology Yearbook 3* (pp. 219-254). Cambridge: Cambridge University Press.

Browman, C., & Goldstein, L. (in press). Tiers in articulatory phonology with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology, I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press.

Bynon, T. (1977). *Historical linguistics*. Cambridge: Cambridge University Press.

Clements, G. N., & Ezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hust and N. Smith (Eds.), *The structure of phonological representations (Part II)* (pp. 213-256). Dordrecht: Foris.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics, 1*, 239-248.

Darwin, C., & Sutherland, N. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic. *Quarterly Journal of Experimental Psychology, 36A*, 193-208.

Dorman, M., Studdert-Kennedy, M., & Raphael, L. (1977). Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics, 22*, 109-122.

Edwards, J., Beckman, M., & Fujimura, O. (1984, May). Contributions of the mandible to lower lip and tongue blade movement. Paper presented at the 107th meeting of the Acoustical Society of America, Norfolk, Virginia.

Eimas, P., & Miller, J. (1980). Contextual effects in infant speech perception. *Science, 209*, 1140-1141.

Fant, G. (1973). Stops in C-V syllables. In G. Fant (Ed.), *Speech sounds and features* (pp. 110-142). Cambridge, MA: MIT Press.

Fant, G., & Lindblom, B. (1961). Studies in minimal speech and sound units. *Speech Transmission Laboratory, Quarterly Progress Report, 2*, 1-11.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception and Psychophysics, 27*, 343-350.

Folkins, J., & Brown, C. (1987). Upper lip, lower lip and jaw interactions during speech. Comments on evidence from repetition-to-repetition variability. *Journal of the Acoustical Society of America, 82*, 1919-1924.

Fowler, C. (1977).*Timing control in speech production*. Bloomington: Indiana University Linguistics Club.

Fowler, C. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica, 38*, 35-50.

Fowler, C. (1983a). Realism and unrealism: A reply. *Journal of Phonetics, 11*, 303-322.

Fowler, C. (1983b). Converging sources of evidence for spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General, 112*, 386-412.

Fowler, C. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics, 36*, 359-368.

Fowler, C., Rubin, P., Remez, R., & Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production, I* (pp. 373-420). London: Academic Press.

Fowler, C., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-135). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fromkin, V. (Ed.). (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Garnes, S., & Bond, Z. (1980). A slip of the ear: A snip of the year? In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 231-240) New York: Academic Press.

Gelfer, C. (1987). *A simultaneous physiological and acoustic study of fundamental frequency declination*. Doctoral dissertation, City University of New York.

Gelfer, C., Harris, K., & Baer, T. (1987). Controlled variables in sentence intonation. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration* (pp. 422-435). Boston: College-Hill Press.

Gentner, D. (1983). Keystroke timing in transcription typing. In W. Cooper (Eds.), *Cognitive aspects of skilled typewriting* (pp. 95-120). New York: Springer-Verlag.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin, .

Green, K., & Miller, J. (1985). On the role of visual rate information in phonetic perception. *Perception and Psychophysics*, 38, 269-276.

Grundt, A. (1976). *Compensation in phonology: Open syllable lengthening*. Bloomington: Indiana University Linguistics Club.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*, 613-617.

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory, JT-8*, 155-159.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics, 4*, 353-363.

Hammarberg, R. (1982). On redefining coarticulation. *Journal of Phonetics, 10*, 123-137.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry, 20*, 253-306.

Hayes, B. (1987, June). Compensatory lengthening in moraic phonology. Paper presented at the First Conference on Laboratory Phonology, Ohio State University, June 5-8, 1987).

Hockett, C. (1955). *Manual of phonology*. (Publications in Anthropology and Linguistics, No. 11). Bloomington: Indiana University Press.

Hombert, J.-M., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language, 55*, 37-58.

Honda, K. (1981). Relationship between pitch control and vowel articulation. In D. Bless & J. Abbs (Eds.), *Vocal-fold physiology* (pp. 286-297). San Diego: College-Hill Press.

Hyman, L. (1973). Consonant types and tone. *Southern California Occasional Papers in Linguistics, 1*, 152-179.

Ingria, R. (1980). Compensatory lengthening as a metrical phenomenon. *Linguistic Inquiry, 11*, 465-495.

Isenberg, D., & Liberman, A. (1978). Speech and nonspeech percepts for the same sound. *Journal of the Acoustical Society of America, , 64*(S1), S20.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press. (1972, Tenth Edition).

Johansson, G. (1973). Visual perception of biological motion. *Perception and Psychophysics, 14*, 201-211.

Johansson, G. (1974). Projective transformations as determining visual space perception. In R. MacLeod & H. Pick (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 117-140). Ithaca: Cornell University Press.

Keating, P. (in press). Mechanisms of coarticulation: Articulatory evidence. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology, I.: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 73*, 322- 335.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology, 15*, 166-171.

Kuhn, G. (1975). On the front-cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America, 58*, 428-433.

Kuhn, G. (1979). Stop consonant place perception with single formant stimuli: Evidence for the role of the front-cavity resonance. *Journal of the Acoustical Society of America, 65*, 774-788.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56*, 485-502.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal-tract shapes from formant frequencies. *Journal of the Acoustical Society of America, 64*, 1027-1035.

Lahiri, A., Gewirth, L., & Blumstein, S. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants. *Journal of the Acoustical Society of America, 76*, 391-404.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA: MIT Press.

Liberman, A. (1982). On finding that speech is special. *American Psychologists, 37*, 148-167.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A., Delattre, P., & Cooper, F. (1952). The role of selected stimulus variables in the perception of the unvoiced-stop consonants. *American Journal of Psychology, 65*, 497-516.

Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157-233). Cambridge, MA: MIT Press.

Lindblom, B., Lubker, T., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming as predictive simulation. *Journal of Phonetics, 7*, 147-161.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm, 21*, 1-59.

Löfqvist, A. (1980). Interarticulatory programming in stop production. *Journal of Phonetics, 8*, 475-490.

Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Swedish obstruents. *Journal of the Acoustical Society of America, 68*, 792-801.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception. *Perception and Psychophysics, 24*, 253-257.

Maeda, S. (1976). *A characterization of American English intonation*. Doctoral dissertation, MIT.

Mann, V. (1980). Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics, 28*, 407-412.

Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition, 24*, 169-196.

Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, J., Aibel, I., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception and Psychophysics, 35*, 5-15.

Miller, J., & Elmas, P. (1983). Studies on the categorization of speech by infants. *Cognition, 13*, 135-165.

Miller, J., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41*, 215-225.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*, 457-564.

Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica, 23*, 65-93.

Nooteboom, S. (1973). The perceptual reality of some prosodic durations. *Journal of Phonetics, 1*, 25-45.

Nooteboom, S., Brokx, J., & DeRooij, J. (1978). Contributions of prosody to speech perception. In W. Levelt & G. D'Arçais (Eds.), *Studies in the perception of language* (pp. 75-107). Chichester: Wiley.

Ohala, J. (1974). Experimental historical phonology. In J. Anderson & C. Jones (Eds.), *Historical Linguistics II* (pp. 353-390). Amsterdam: North Holland.

Ohala, J. (1978). Production of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5-39). New York: Academic Press.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. Hendrick, & M. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). *Chicago: Chicago Linguistics Association.*

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75*, 224-230.

Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measures. *Journal of the Acoustical Society of America, 39*, 151-168.

Pagliuca, W. (1982). *Prolegomena to a theory of articulatory evolution*, Doctoral dissertation, State University of New York, Buffalo.

Pierrehumbert, J. (1979). The perception of fundamental frequency. *Journal of the Acoustical Society of America, 66*, 363-369.

Pierrehumbert, J. 1980). *The phonology and phonetics of English Intonation*. Doctoral dissertation, MIT.

Porter, R. (1978). Rapid shadowing of syllables: evidence for symmetry of speech perceptual and motor systems. Paper presented at the meeting of the Psychonomic Society, San Antonio.

Porter, R., & Lubker, J. (1980). Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage in speech. *Journal of Speech and Hearing Research, 23*, 593-602.

Saltzman, E. (1986). Task-dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 129-144). (Experimental Brain Research Series 15). New York: Springer-Verlag.

Shattuck-Hufnagel, S., & Klatt, D. (1979). Minimal use of features and markedness in speech production. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Silverman, K. (1985). Vowel intrinsic pitch influences the perception of intonational prominence. *Journal of the Acoustical Society of America, 77*, S38.

Silverman, K.)1986). F_0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica, 43*, 76-91.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Doctoral dissertation, Cambridge University.

Stevens, K. (1960). Toward a model for speech recognition. *Journal of the Acoustical Society of America, 32*, 47-55.

Stevens, K., & Blumstein, S. (1981). The search for invariant correlates of phonetic features. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, K., & House, A. (1956). Studies of formant transitions using a vocal-tract analog. *Journal of the Acoustical Society of America, 28*, 578-585.

Studdert-Kennedy, M., & Shankweiler, D. 1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America, 48*, 579-594.

Walley, A., & Carrell, T. (1983). Onset spectra and formant transitions in the adult and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 73*, 1011-1022.

Wang, W. (1972). The many uses of F_0. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 487-504). The Hague: Mouton.

Weiss, P. (1969). The living system: Determinism stratified. In A. Koestler & J. Smythies (Eds.), *Beyond reductionism* (pp. 3-55). Boston: Beacon Press.

Whalen, D. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. *Journal of the Acoustical Society of America, 69*, 275-282.

Whalen, D. (1984). Subcategorical mismatches slow phonetic judgments. *Perception and Psychophysics, 35*, 49-64.

Whalen, D., & Liberman, A. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237*, 169-171.

FOOTNOTES

[*] To appear in T. Tighe, B. Moore, & J. Santroch (Eds.), *Human development and communication sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates. (in press).

[†] Also Dartmouth College.

[1] Formants are resonances of the vocal tract. Transitions are changes in vocal-tract resonances due to vocal-tract gestures.

[2] Place of articulation" refers to the place in the vocal tract of a constriction for a consonant. /b/ and /p/ are produced with a constriction at the lips; /t/ and /d/ have a constriction at the alveolar ridge of the palate behind the teeth; /g/ and /k/ have their constriction at the soft palate.

[3] Phonetic features are attributes of consonants and vowels. The collection of features for a phonetic segment identify it uniquely.

[4] Stop consonants are phonetic segments produced with a complete constriction of the vocal tract. In English, they are /b, p, d, t, k, g/. Short-time spectra are frequency-by-amplitude profiles of the stops averaged over the first 20-30 ms after release of the consonantal constriction.

[5] Browman and Goldstein (e.g., 1986, in press) are developing an articulatory phonology in which abstract phonetic gestures will do the work of phonetic features in other phonologies. Accordingly, the primitives of their theory may better fit the characteristics of vocal-tract activity than the static phonetic features proposed by other theorists.

[6] Ohala does not put it quite this way. In his view, coarticulation distorts the way segments are produced. Systematic mishearings occur when listeners fail to recognize distortion as such.

[7] Functional coupling is demonstrated most convincingly by "perturbation experiments" (see Folkins and Brown (1987) for a review and discussion). For example, Kelso, Tuller, Vatikiotis-Bateson, and Fowler (1984) perturbed the closing gesture of the jaw for the final consonant in /bæb/ or /bæz/ by applying a brake to the jaw. When the jaw brake was applied during

/bæb/, the upper lip responded to the perturbation within 20-30 ms with an increased lowering gesture and with increased activity of the orbicularis oris muscle. The very short latency response implies a low-level coupling between jaw and lips. An additional finding was that the "compensatory" response of the lips occurred only in productions of /b/ in /bæb/, where it is compensatory, not in productions of /z/ in /bæz/, where it is not. Therefore, the coupling is "functional"; it is established just when bilabial closure is required. It is not an enduring reflex of the vocal-tract. The coupling is most likely not achieved just in case someone applies a brake to the jaw. Bilabial consonants produced in the context of open vowels are produced with a more open jaw, and hence with more extensive closing gestures by the lips, than bilabial consonants in the context of closed vowels (Keating, Lindblom, and Lubker, cited in Keating, in press). The coupling of jaw and lips enables the same organization of the vocal tract to achieve closure in the context of different "perturbations" applied to the jaw by the requirements of different coarticulating vowels. (See Saltzman, 1986.)

[8] There are two caveats to the claim that the serial ordering of constriction gestures is strictly maintained. Sometimes in casual or rapid speech, the closing gesture for a consonant may cooccur with that for a later consonant (see Browman and Goldstein, in press). For example, in data described by Browman and Goldstein, a rapid speech version of "perfect memory" is transcribed phonetically as if the /t/ had been deleted by the talker. However, X-ray tracings of tongue tip movement reveal that the alveolar closure is achieved. The reason for the transcription is that the alveolar constriction occurs simultaneously with the more forward closure for the first /m/ in "memory," which therefore largely prevents evidence for /t/ in the acoustic signal. In this case, the serial ordering of the consonants is not maintained in articulation, but then listeners do not hear the ordering either. In the second case, constriction gestures for a vowel may sometimes reach their peak later than that for a following consonant (for example, in the vowel and final consonant of "five" in X-ray data described by Edwards, Beckman, & Fujimura, 1984). In this case , however, the serial ordering of the segments is still preserved, because the opening gesture for the vowel begins well before the closing gesture for the consonant seals off the oral cavity (see Figure 1 in Edwards et al.).

[9] Voice onset time is the time between stop release and onset of voicing for a following voiced segment. It is longer for /p/ than for /b/ and longer for slow /p/ than for fast /p/.

[10] Notice that Silverman's studies of intrinsic f_0 and consonant voicing are complementary in an interesting way. One shows that effects of intrinsic f_0 are factored from an intonational peak. The other shows that effects of an intonational contour are factored from f_0 perturbations due to consonant voicing. Together the findings imply that neither perception of suprasegmental information (specifically of an intonation contour) nor perception of segmental information (to which intrinsic f_0 and perturbations due to voicing contribute) can be temporally primary. That is, prior identification of the intonation contour cannot be used to determine intrinsic f_0, because identification of peak heights in the contour depend on knowing intrinsic f_0. Yet prior identification of segmental f_0 perturbations cannot be used to determine the intonation contours, because the contour is used to determine those perturbations. The factoring is mutual and simultaneous.

[11] Schwa is a reduced vowel, always unstressed, and symbolized /ə/.

[12] /l/ and /r/ are not distinct phonological segments in Japanese as they are in English, and Japanese listeners have difficulty discriminating and identifying them (Miyawake, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975).

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 130-148

An Electropalatographic and Acoustic Study of Consonant-to-Vowel Coarticulation*

Daniel Recasens†

Coarticulatory effects from consonants on vowels are investigated by means of electropalatography and acoustical analysis. The electropalatographic data confirm the validity of an articulatory classification of vowels in a front and a back categories. Moreover, vowels are shown to differ in degree of sensitivity to coarticulatory effects from the adjacent consonants. In general, consonant-dependent effects are larger upon those articulatory regions which do not intervene in the formation of a vowel constriction. Other gestural constraints (e.g., lip rounding for [u]) are called forth to interpret some data in lingual coarticulation. The paper also concludes that the degree of C-to-V coarticulation is influenced by the production requirements on the consonant; that explains why the degree of linguopalatal contact for [i] is more affected by velarized [ɫ] than by other consonants. The electropalatographic data are correlated with acoustic (F2) data.

1. INTRODUCTION

While there is a good number of studies on vowel-to-consonant coarticulation, rather less attention has been paid to the coarticulatory effects of consonants on vowels. A possible rationale for this preference lies in the articulatory constraints used by speakers during the production of the two phonetic categories: on the one hand, vowels are resistant to coarticulation since they are produced by means of global vocal tract shapes and require articulatory control upon the entire tongue body configuration; on the other hand, consonants involve local constrictions and leave articulatory regions free to coarticulate (Fowler, Rubin, Remez, & Turvey, 1980).

Studies on C-to-V coarticulation show that vowels differ in degree of coarticulatory sensitivity. However these studies are almost entirely acoustic (Pols, 1977; Recasens, 1985; Stevens & House, 1963). In the absence of relevant production studies the articulatory correlates of the observed C-to-V effects must be inferred from the acoustic data. Articulatory data (such as those presented in this paper) are needed to gain some more understanding about the production mechanisms used by speakers when uttering CV or VC syllables. They should help to improve current speech production models as well.

A goal of this paper is to show that vowels are subject to coarticulatory effects from consonants in a way which is comparable to V-to-C coarticulatory trends. Thus, it is possible to attribute constriction locations to vowels (Wood, 1979) and generate a model of C-to-V coarticulation according to which some tongue regions are more sensitive to coarticulatory effects than others. The issue is, then, to investigate the size and the nature of the C-to-V effects and to come up with a rationale for such effects which is based on the articulatory properties of the vowels and consonants. An important notion here is that the amount of C-to-V coarticulation for a particular articulator is inversely related to the involvement of that articulator in the formation of the vowel constriction. Accordingly, effects on tongue dorsum activity ought to be much larger for [ə] than for [i].

The analysis of coarticulation presented here is both electropalatographic (EPG) and acoustic (F2). An inconvenience with the palatographic

---

This research has been supported by NINCDS Grant NS-13617 to Haskins Laboratories, the ESPRIT-ACCOR project from the European Community and a fellowship from the Catalan Government (CIRIT). I would like to thank the comments by Carol A. Fowler, Shinji Maeda and Ignatius G. Mattingly on previous versions of this paper.

technique is that it provides accurate linguopalatal impressions for front vowels but little or no information for back vowels. This problem becomes less obvious when, as in the present investigation, back vowels are embedded in a consonantal context. In those circumstances consonant-dependent fronting and raising effects upon lingual activity cause linguopalatal contact during the production of the vowel; it thus becomes possible to analyze some C-to-V effects. The relationship between the EPG and the acoustic data is also studied in this paper. Only F2 data will be analyzed in view of the fact that there is a main correlation between the values of this formant, and the degree and location of linguopalatal contact (Fant, 1960).

2. METHOD

The sequences of the speech material conform to the phonology of Catalan. They share the segmental structure [ˈCVCə] when V1 is unreduced and [CVˈCa] when V1 is reduced. The V1 slot in the [ˈCVCə] sequences was filled with all the unreduced Catalan vowels, namely, [i], [e], [ɛ], [a], [ɔ], [o] and [u]; only [ə] was uttered as V1 in the [CVˈCa] sequences. In all sequences C1 was identical to C2. The following Catalan consonants were chosen in order to analyze a large variety of C-to-V effects: bilabial [p]; apicodental [t]; apicoalveolar or laminoalveolar [s]; lamino-postalveolar [ʃ]; predorso-alveolopalatal [ɲ]; dorsopalatal [j]; front or back dorsovelar [k]; velarized apicoalveolar [ɫ].

One Catalan speaker (the author) uttered the sequences ten times in the carrier sentence *Diu* ___ ("He says ___"). The overall number of sequences was 640 (8 vowels x 8 consonants x 10 repetitions). Simultaneous electropalatographic and acoustic recordings were made. The electropalatographic system (Rion Electropalatograph model DP-01) is equipped with 63 equidistant electrodes arranged in five semicircular rows (see Figure 1) and enables the recording of one pattern of contact every 15.6 ms. The distance between adjacent electrodes on the artificial palate used in the experiment reported here varies between 5 and 6 mm. Electrodes have been given a code number for each side of the palate, starting from the backmost electrode up to the frontmost electrode; thus, there are 8 electrodes on row 1, 7 on row 2, etc. It can be seen that the number of electrodes decreases as the rows become more central. Rows have also been numbered from 1 (peripheralmost row) through 5 (centralmost row). The figure also shows that some electrodes fall along a median line and, more specifically, that the two backmost electrodes on that line are not associated with any of the five semicircular rows. A division in articulatory zones (alveolar, prepalatal, mediopalatal, postpalatal) has been superimposed upon the surface of the electropalate.

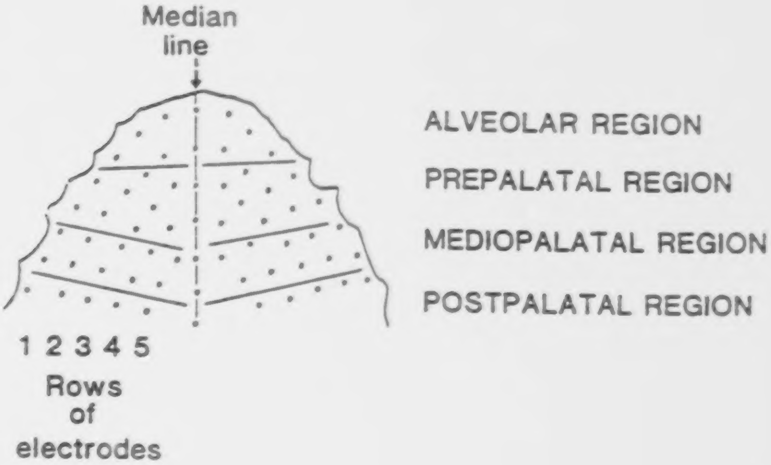


Figure 1. Electropalate.

The EPG frame corresponding to the vowel midpoint (as detected on waveform displays) was selected for measurement. A valid criterion was the averaging of the number of "on" electrodes on each row across repetitions. Each average number corresponds to the point of maximal fronting since all electrodes behind the frontmost electrode on each row were always "on" for all the vowels in all the consonantal contexts. Three F2 measurements were taken at the vowel midpoint by means of an LPC analysis program available in the ILS (Interactive Laboratory System) package; F2 data presented in this paper represent averages across repetitions.

3. ELECTROPALATOGRAPHIC DATA

3.1 Vowel-dependent linguopalatal configurations

3.1.1 Articulatory characteristics

Figure 2 displays linguopalatal contact for each vowel in the context *pVp*. Since the bilabial stop involves no active tongue movement during its production it should have a minimal effect upon the tongue movement of the adjacent vowel.

Electrodes have been coded on the vertical axis; rows have been coded on the horizontal axis. The contour lines in the figure connect the points of maximum linguopalatal fronting on each row. Linguopalatal contact takes place over the area between the contour lines and the sides of the figure. The vertical line along the median line of electrodes separates patterns of linguopalatal contact at the right and left sides of the artificial palate; the division in articulatory zones coincides with that in Figure 1.

The place of maximum constriction for the front vowels ([i], [e] and [ɛ]) in the figure is towards the mediopalate and postpalate and involves the tongue dorsum. The figure also shows a decrease of contact area all over the palatal surface in the progression [i]>[e]>[ɛ], which is in agreement with X-ray data for several languages (Wood, 1982). These three front vowels differ in the degree of contact along the front/back and right/left dimensions: the tongue sides achieve the alveolar zone for [i], hardly so for [e] and not at all for [ɛ]; central contact decreases in the same progression (i.e., [i]>[e]>[ɛ]). Therefore, tongue fronting decreases with decreasing tongue height. These EPG data are consistent with the view that the fronting and raising of the tongue body for the production of front vowels are executed by means of a single gesture (also Harshman, Ladefoged, & Goldstein, 1977).

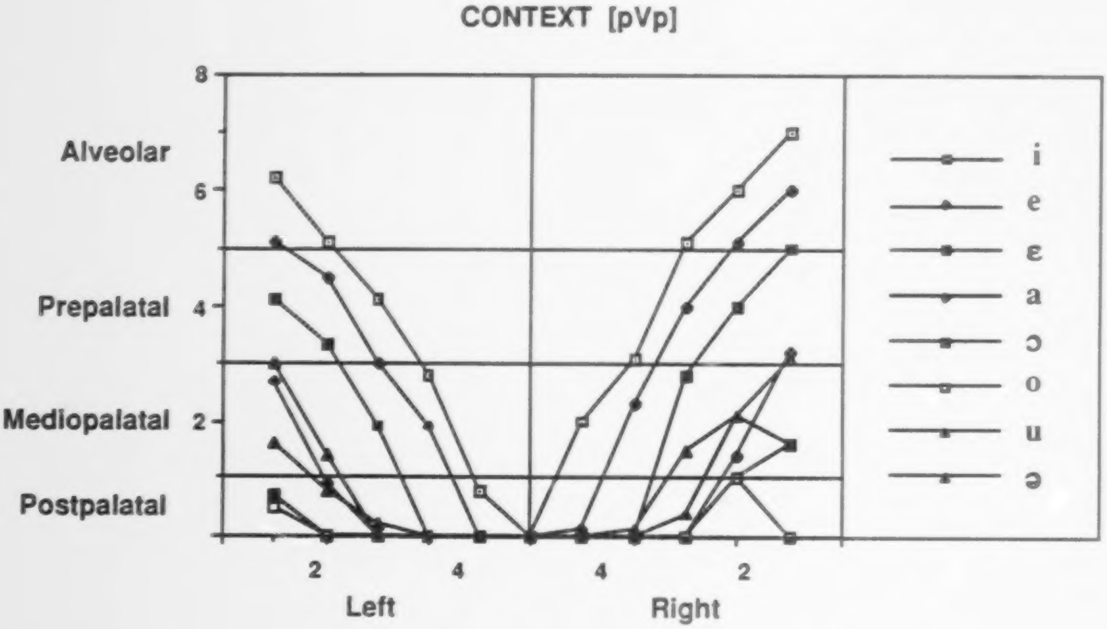


Figure 2. Linguopalatal contact for each vowel in the context [pVp].

The vowels [a], [ɔ], [o] and [u] are articulated farther back on the palatal surface; indeed, frontmost contact in this case hardly reaches the prepalate. Those are therefore back vowels. It can also be seen that the linguopalatal traces for [ɔ] and [o] lie behind those for [a] and [u]. Thus it appears that vowels which are articulated with an upper pharyngeal constriction (i.e., [ɔ] and [o]) show less palatal contact than those that are articulated either at the lower pharynx (i.e., [a]) or at the velar region (i.e., [u]). A rationale for this contrast is to be found in the overall tongue body configurations (see, for example, X-ray diagrams in Fant, 1960): the front dorsum is lowered and retracted during the production of [u], [ɔ] and [o], and lowered but not retracted during the production of [a]. As a result of these mechanisms the front dorsum for [a] is flat and more fronted than for back rounded vowels. Larger front dorsum contact for [u] than for [ɔ] and [o] is consistent with a higher position of the jaw and tongue body.

The linguopalatal configuration for the schwa is more similar to that for back vowels (particularly [a]) than to that for front vowels. This is so because no raising of the front dorsum is required for the production of [ə], particularly when the schwa is coarticulated with a low vowel (i.e., V2=[a]).

These data show that there is a good correlation between the EPG traces, and a front and a back class of vowels. Recent production studies using other analysis techniques confirm the validity of this articulatory division (Alfonso & Baer, 1982; Baer, Alfonso, & Honda, 1988).

3.1.2 Variability of linguopalatal contact and linguopalatal asymmetry

Figures 3 through 10 show linguopalatal contact in number of "on" electrodes for all the Catalan vowels as a function of all the adjacent consonants. Moreover Table 1 gives the means and standard deviations (SD) for all the vowels across consonantal contexts on each row of electrodes, both at the left and at the right sides of the palate. Empty slots in the table represent mean values smaller than .5. SD values give an estimate of the variability in degree of linguopalatal fronting on each row.

The means show the existence of asymmetry in degree of palatal contact for all vowels on all rows. As it can also be seen in the figures, the number of "on" electrodes on the right side of the palate is always larger than that on the left side. These data indicate that lingual asymmetry does not only occur for consonants (Hamlet, Bunnell, & Struntz, 1986; Marchal & Espesser, 1989) but also for vowels.

Table 1. *Means and SD values in number of "on" electrodes across consonantal contexts. Data are given for each row of electrodes (1 through 5) on the right (RS) and left (LS) sides of the palate.*

		Row 1		Row 2		Row 3		Row 4		Row 5	
		X	SD	X	SD	X	SD	X	SD	X	SD
[i]	LS	5.8	.5	4.8	.5	3.8	.4	2.4	.5	.5	.6
	RS	6.4	.4	5.8	.7	5.1	.7	3.1	.5	1.8	.7
[e]	LS	5	.6	4.3	.6	3	.6	1.4	.7	—	—
	RS	5.8	.5	5.2	.7	4.1	.7	2.3	.9	—	—
[ɛ]	LS	4.1	.4	3.2	.5	1.7	.7	—	—	—	—
	RS	5	.5	4	.6	2.9	.9	.7	.7	—	—
[a]	LS	3.3	1	1.7	1.2	—	—	—	—	—	—
	RS	3.8	1.3	2.4	1.2	1	1.2	—	—	—	—
[ɔ]	LS	1.9	1.5	.6	.9	—	—	—	—	—	—
	RS	2.7	1.7	1.4	.6	—	—	—	—	—	—
[o]	LS	2.3	1.8	1.2	1.4	—	—	—	—	—	—
	RS	2.5	2.2	2	1.4	.6	.9	—	—	—	—
[u]	LS	3.6	1.8	2.6	1.9	1.7	1.3	—	—	—	—
	RS	3.7	2.7	3.6	1.9	2.5	1.8	1	1.2	—	—
[ə]	LS	4.8	1.4	3.4	2	2	1.5	.7	1	—	—
	RS	5.4	1.7	4.7	2	3	2.1	1.1	1.5	—	—

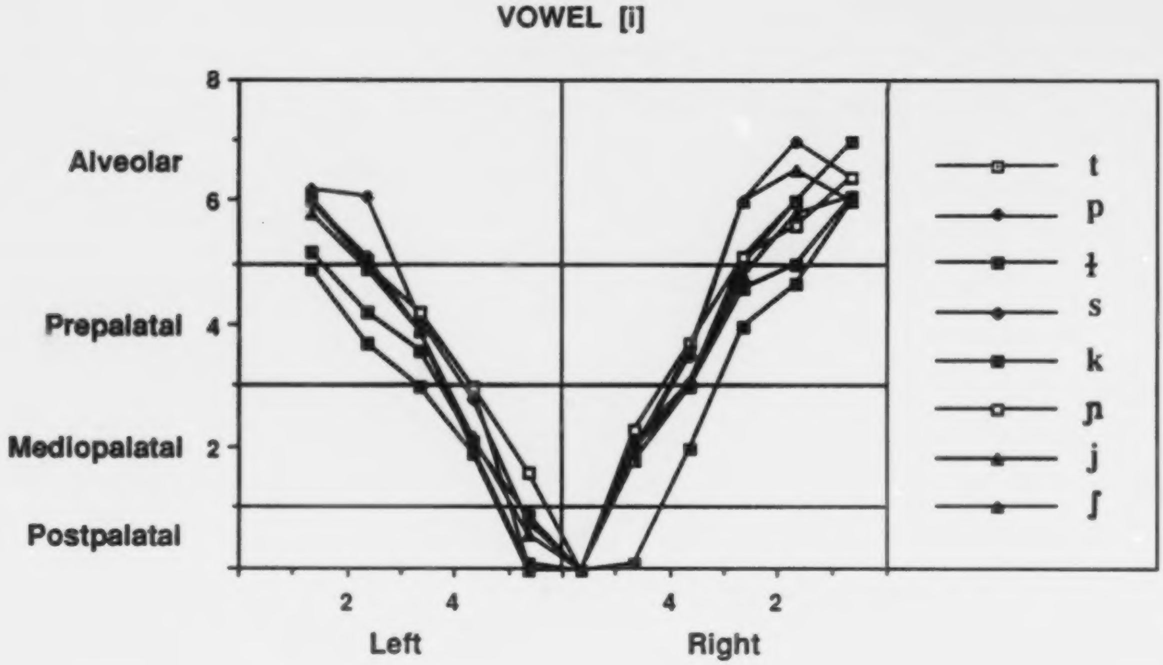


Figure 3. Linguopalatal configuration for [i] as a function of consonantal environment.

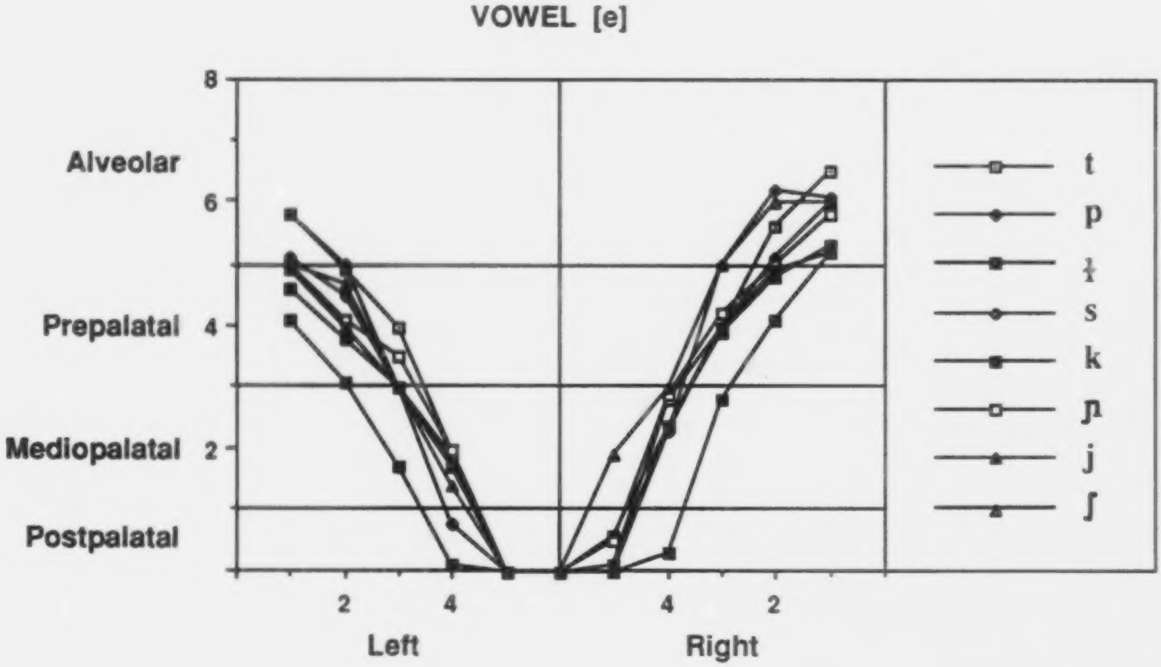


Figure 4. Linguopalatal configuration for [e] as a function of consonantal environment.

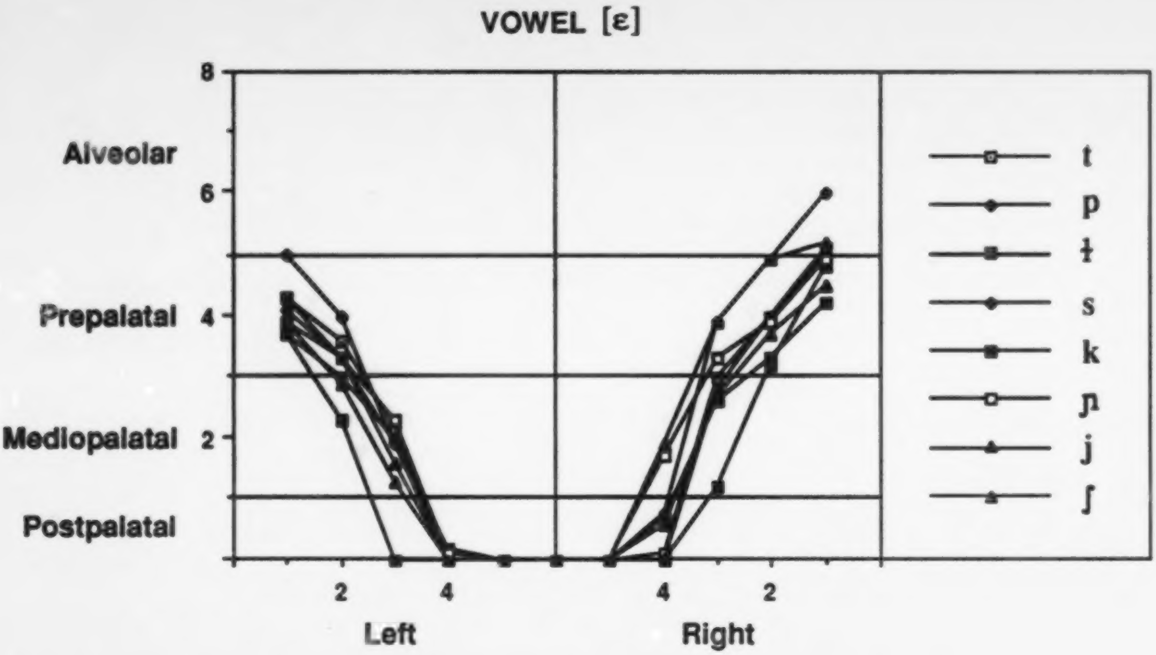


Figure 5. Linguopalatal configuration for [ɛ] as a function of consonantal environment.

The SD values allow grouping the vowels into two categories, front and back. The group of front vowels (i.e., [i], [e], [ɛ]) always shows SD values below 1; Figures 3, 4 and 5 show indeed little variability in the configuration of the contour lines as a function of consonantal context. The group of back vowels (i.e., [a], [ɔ], [o], [u]), on the other hand, often shows SD values above 1; Figures 6, 7, 8 and 9 show large degrees of linguopalatal variability across consonantal environments.

In the case of front vowels, the fact that SD values are equally low on all rows suggests that the entire tongue dorsum is subject to a high degree of constraint. This finding is consistent with current hypotheses about the production of [i]. X-ray data reported by Perkell and Nelson (1985) show similar contextual variability for pellets positioned at the blade and the mediodorsum; small variability in vertical displacement suggests that those two lingual regions are positioned as a unit during the articulation of the vowel [i]. Kiritani et al. (1977) also found less context-dependent variability at the blade and mediodorsum than further back in the tongue for the Japanese vowels [i] and [e].

A high degree of contact variability for back vowels is predictable from their constriction location. In all cases the front dorsum is left quite free to coarticulate with the contextual consonants since the vowel constriction is formed with the back of the tongue. Consistently, Kiritani et al. (1977) show larger context-dependent effects at the blade and mediodorsum than further back in the tongue for the Japanese vowels [a], [o] and [u]. Moreover SD values are higher for [u] than for [a], [ɔ] and [o] (see Figures 6 through 9. Also Recasens, 1985; Stevens & House, 1963). A rationale for the large degree of variability exhibited by [u] is associated presumably with the lip rounding gesture (see § 3.2.3).

SD values are also very high for the schwa. Indeed, Figure 10 shows a great deal of consonant-dependent variability in linguopalatal contact. This is an expected trend since [ə] is unconstricted and should be highly sensitive to changes in the vocal tract configuration induced by the surrounding consonants.

3.2 Consonant-dependent effects in linguopalatal contact

3.2.1 General characteristics

Figures 3 through 10 indicate that the size of the SD values given in Table 1 depends on the articulatory relationship between particular consonants and particular vowels. In order to find out more about the precise effect of consonants on vowels, the following procedure was applied.

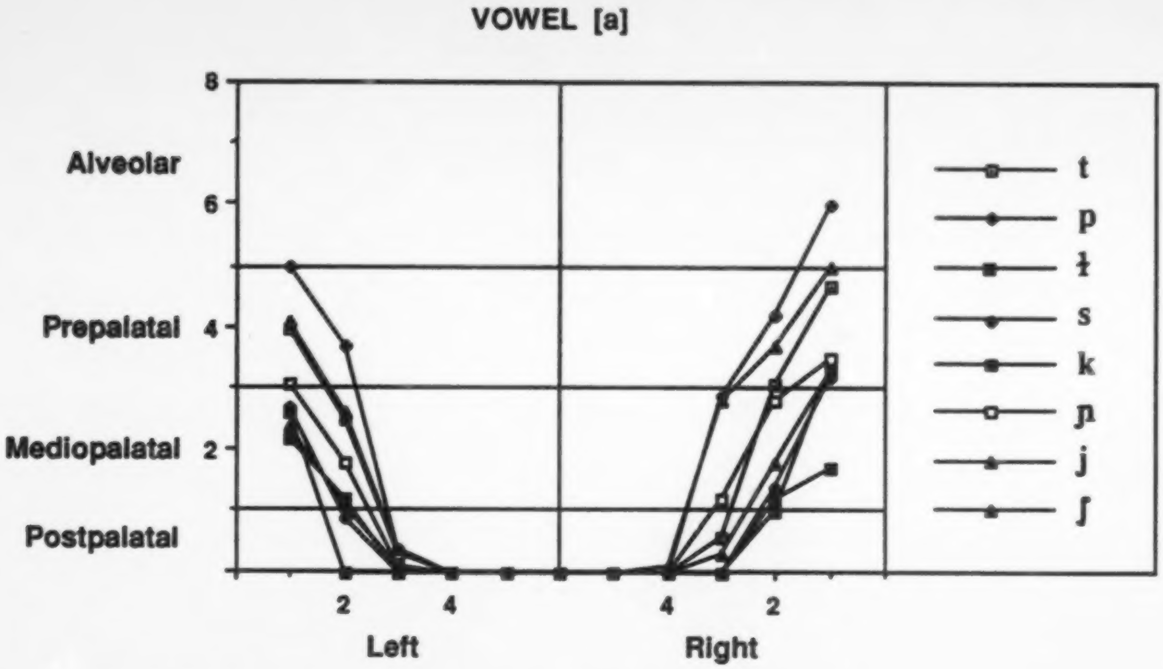


Figure 6. **Linguopalatal configuration for [a] as a function of consonantal environment.**

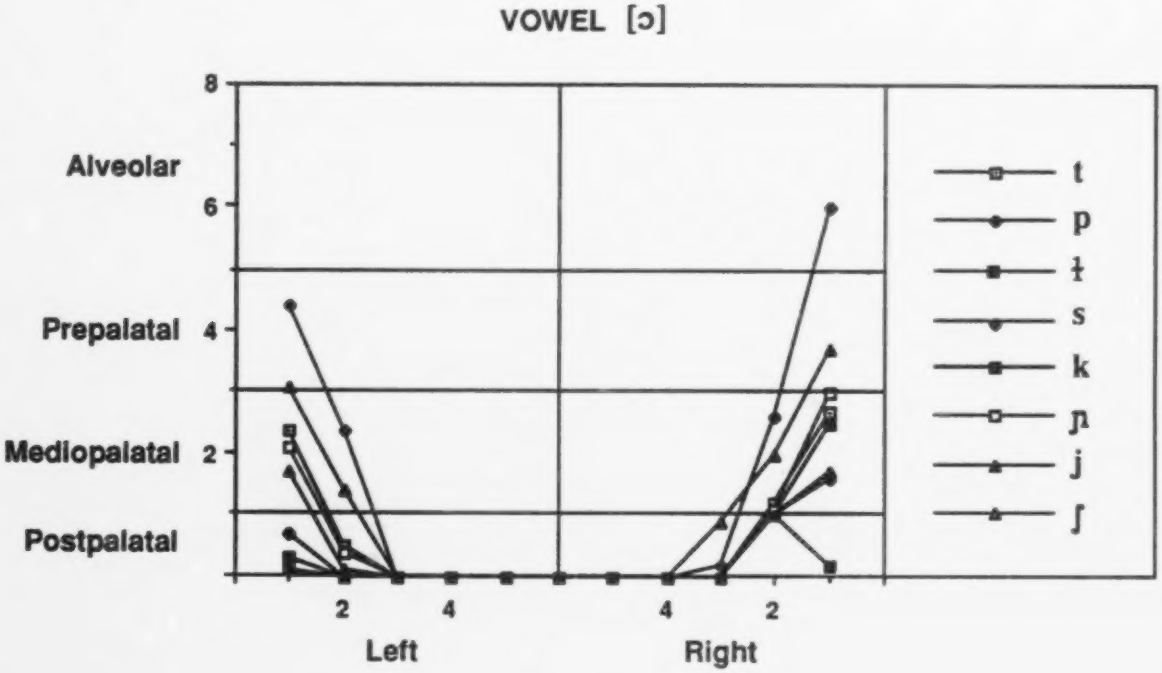


Figure 7. **Linguopalatal configuration for [ɔ] as a function of consonantal environment.**

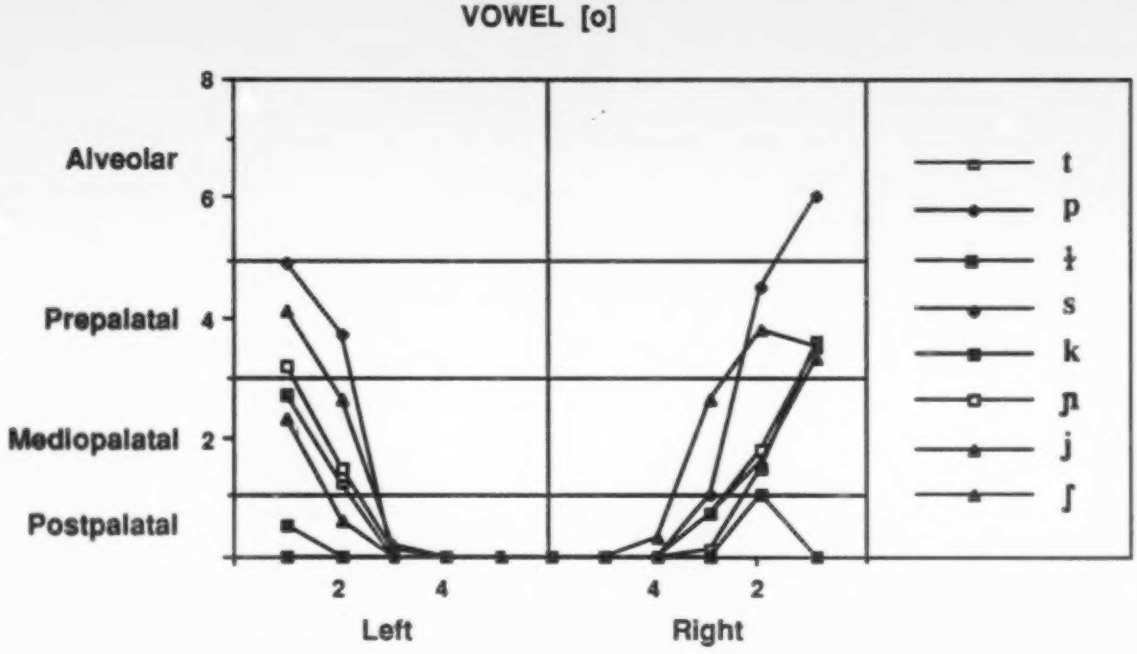

VOWEL [o]

Figure 8. Linguopalatal configuration for [o] as a function of consonantal environment.

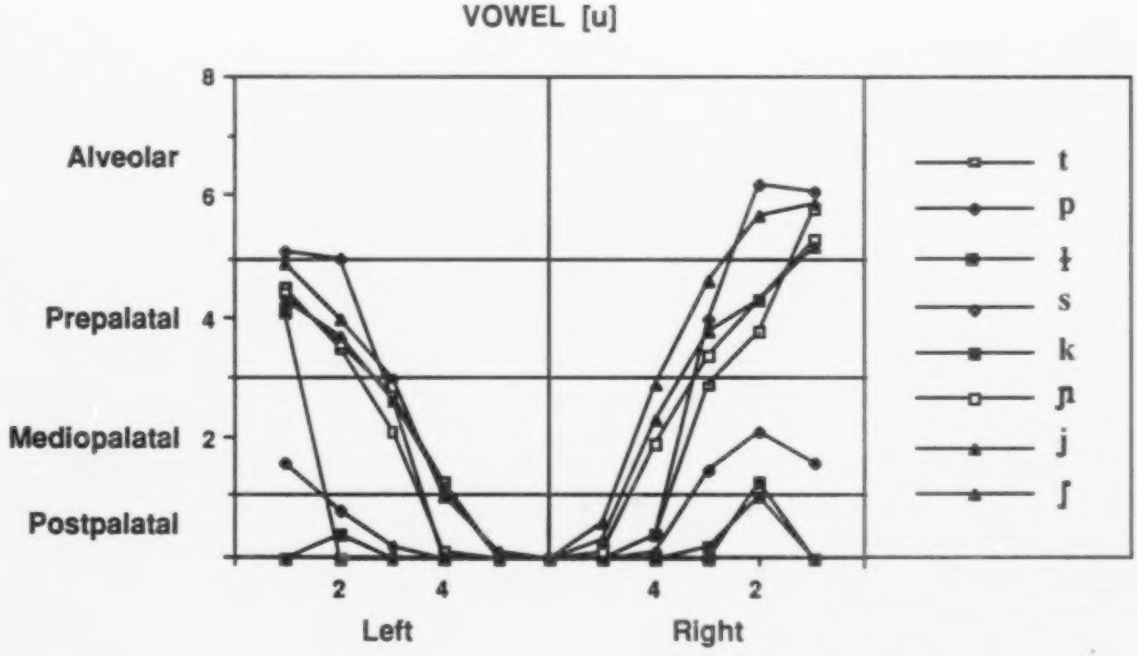

VOWEL [u]

Figure 9. Linguopalatal configuration for [u] as a function of consonantal environment.

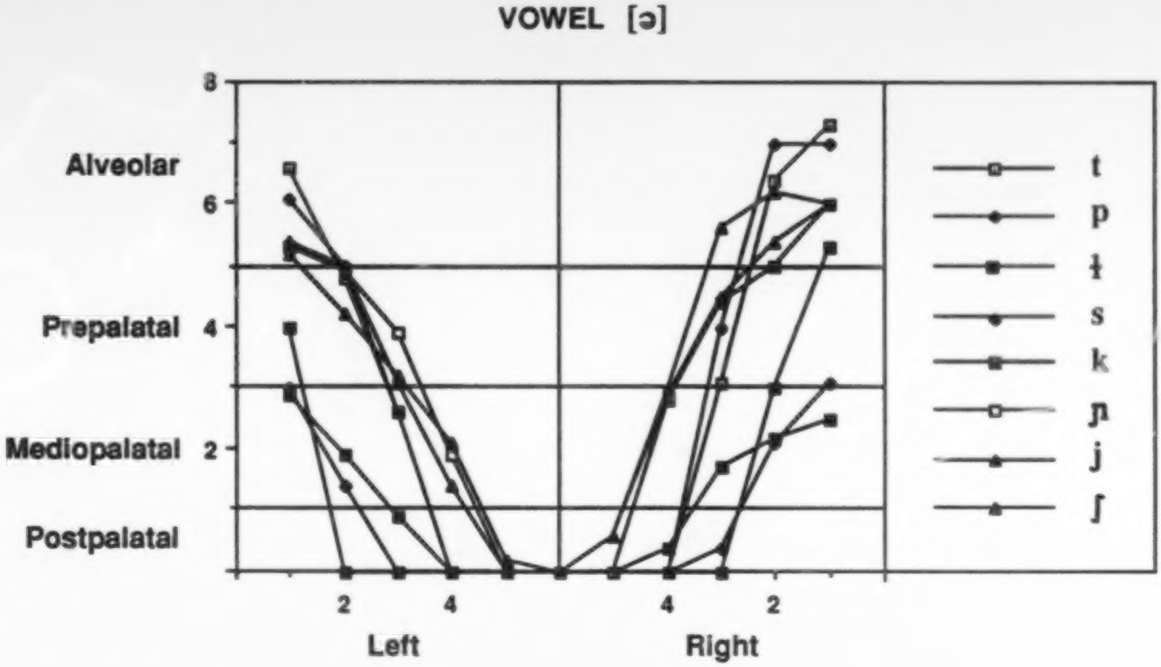


Figure 10. Linguopalatal configuration for [ə] as a function of consonantal environment.

In the first place, an index of linguopalatal fronting was obtained for each vowel in each consonantal context; for that purpose, the average number of "on" electrodes was calculated for each row across the right and left sides of the palate. Secondly, each index of linguopalatal fronting was subtracted from the fronting index in the neutral context *p*V*p*. Results are plotted on Figures 11 through 18. The figures show the amount of deviation in number of electrodes for each vowel with respect to the neutral *p*V*p* context as a function of row number and consonantal context. Positive effects reflect an increase in linguopalatal contact with respect to the linguopalatal configuration in the *p*V*p* context; negative effects reflect a decrease in contact with respect to the neutral configuration. The magnitude of the deviation can be interpreted as a measure of the degree of C-to-V coarticulation.

In the case of the front vowels [i], [e] and [ɛ] (Figures 11, 12 and 13) the number and size of the positive effects (i.e., effects in tongue dorsum raising and/or fronting) are not too different from those corresponding to the negative effects (i.e., effects in tongue dorsum lowering and/or backing). Moreover the contribution of the negative effects decreases as the vowel becomes lower while the contribution of the positive effects increases in the same progression. A low palatal vowel can become higher when adjacent to consonants which are produced with a higher tongue dorsum position; effects on a high palatal vowel, on the other hand, are mainly lowering effects.

Back vowels [a], [ɔ], [o] and [u] (Figures 14 through 17) reveal a different pattern of C-to-V coarticulation. Firstly, there are almost no negative effects in this case: while velar [k] and velarized alveolar [ɫ] mainly cause negative effects, all the remaining consonants cause an increase in degree of linguopalatal contact. As discussed in § 3.2.3, these coarticulatory effects accord well with the low positioning of the front dorsum of the tongue for back vowels, and the placement of that tongue region for the adjacent consonants (i.e., low in the case of [k] and velarized [ɫ], and quite high in the case of dentoalveolars and palatals). Effects on the peripheral rows are larger for the higher vowels [o] and [u] than for the lower vowels [ɔ] and [a]; they reflect tongue fronting and may very well be associated with a considerable lip rounding gesture during the vowel. Effects on the central rows 4 and 5 are only found for [u] consistently with the fact that fronting the tongue body during a high back vowel should involve an increase in dorsopalatal contact.

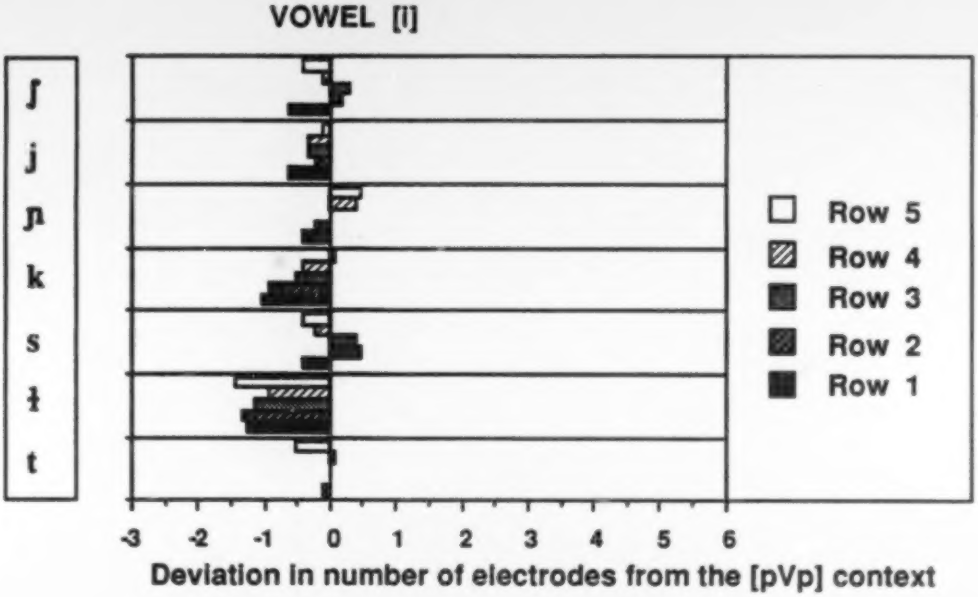


Figure 11. **Amount of deviation in number of electrodes for [i] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

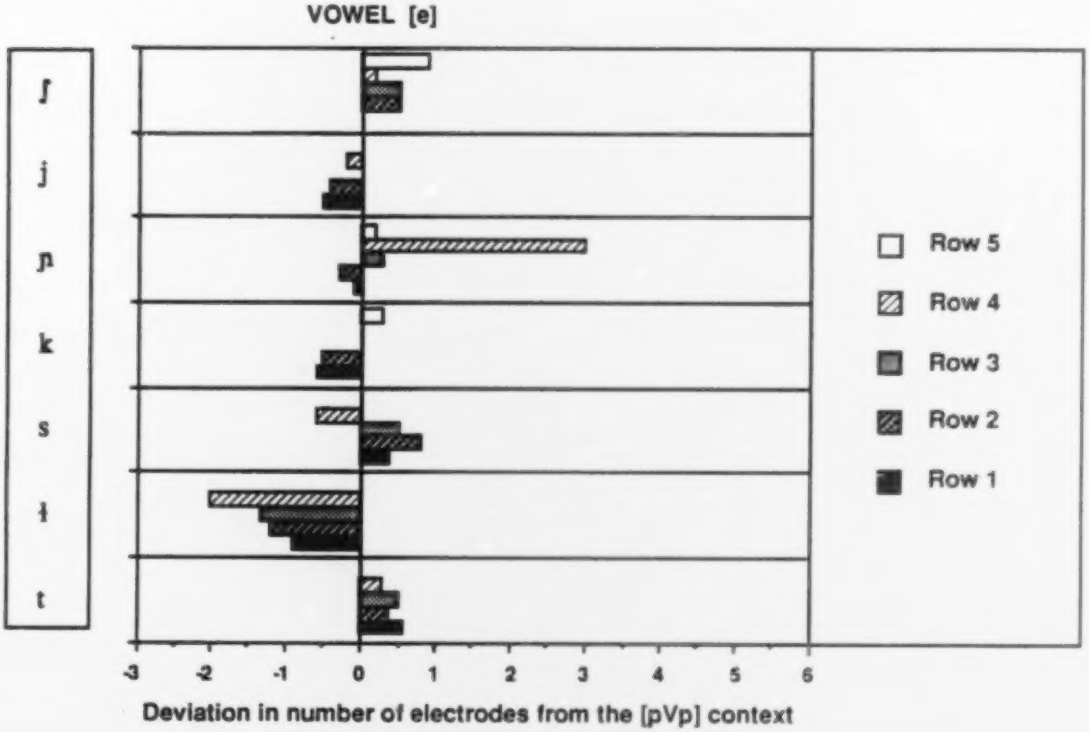


Figure 12. **Amount of deviation in number of electrodes for [e] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

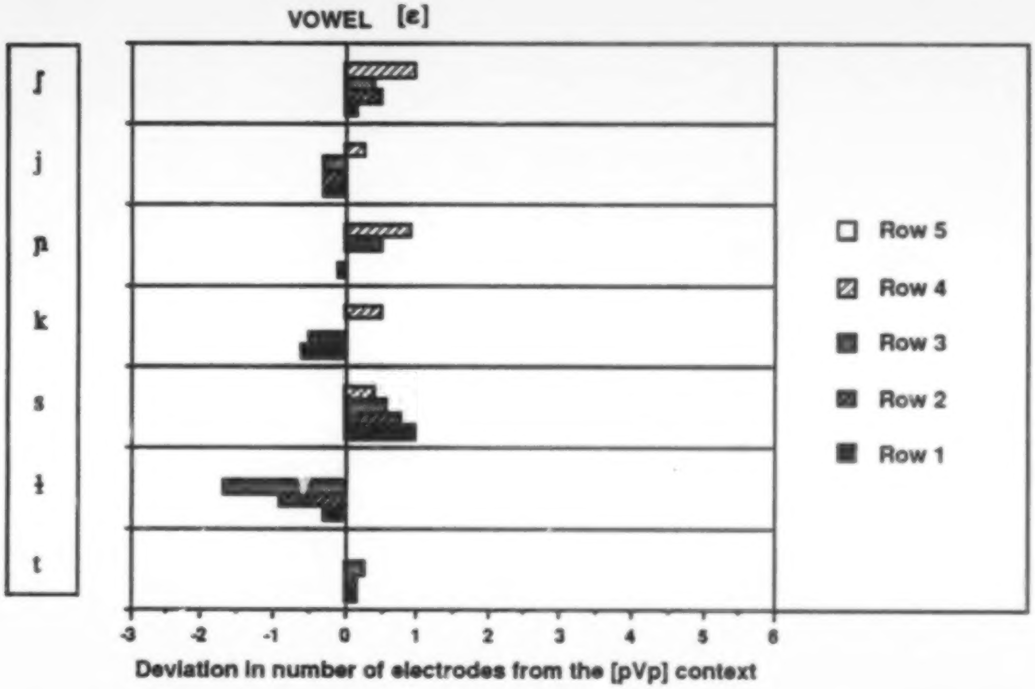


Figure 13. **Amount of deviation in number of electrodes for [ɛ] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

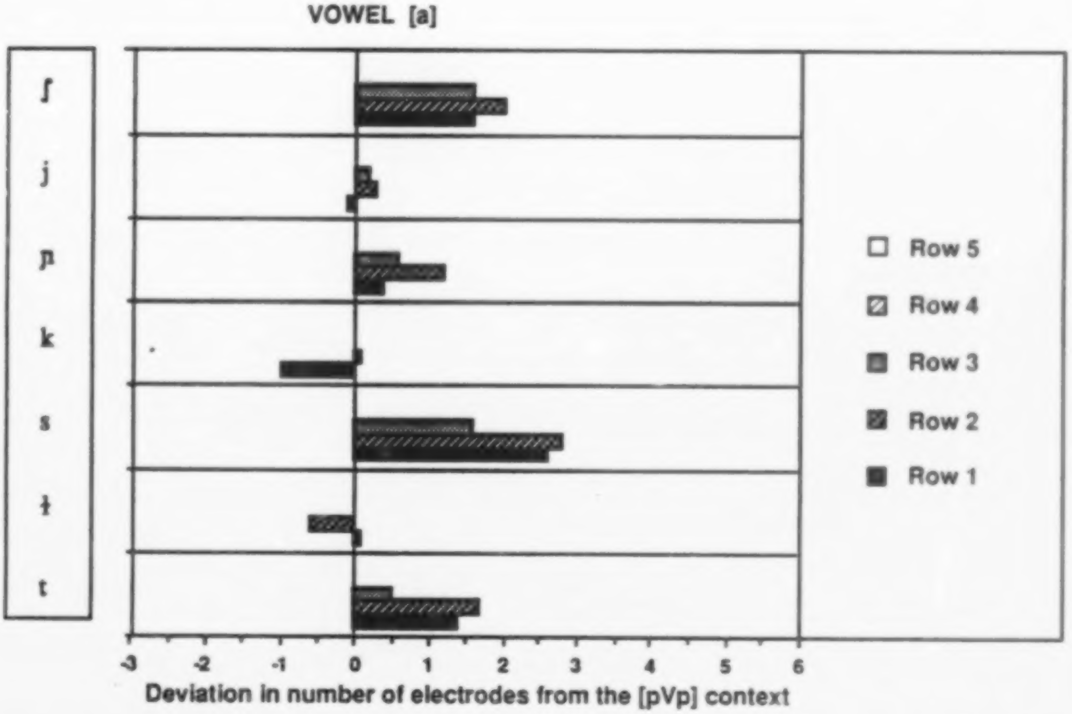


Figure 14. **Amount of deviation in number of electrodes for [a] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

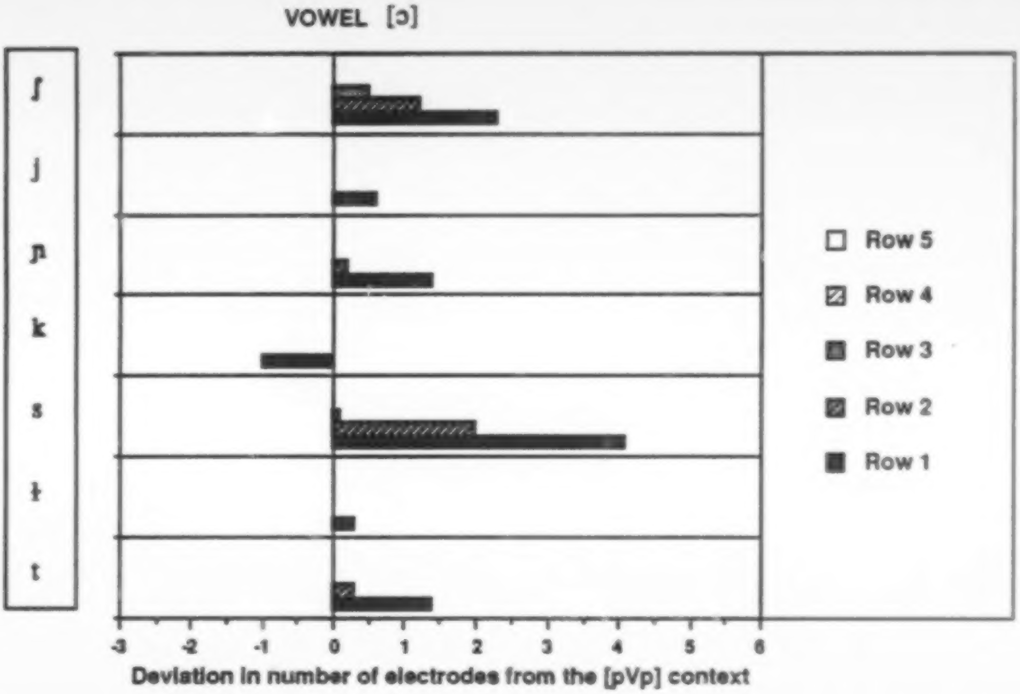


Figure 15. **Amount of deviation in number of electrodes for [ɔ] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

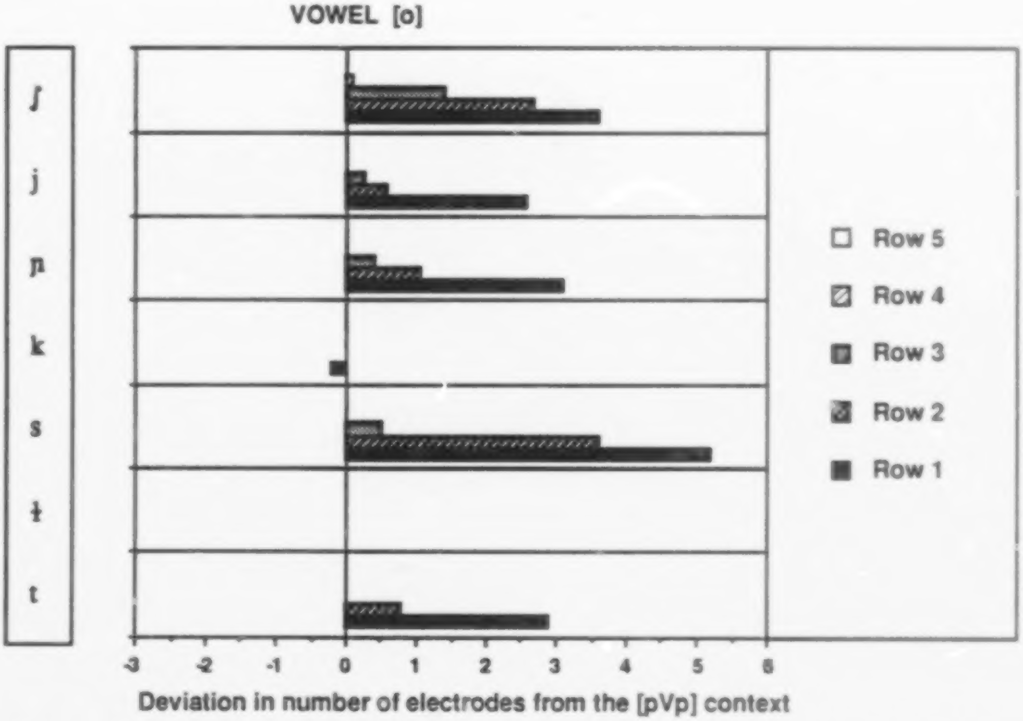


Figure 16. **Amount of deviation in number of electrodes for [o] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

The schwa (Figure 18) shows a similar pattern to that exhibited by [u] except for the rarity of negative effects and the presence of large positive effects at the center of the palatal surface. The front dorsum is thus more sensitive to coarticulation for the schwa than for back vowels. That tongue region is subject to less articulatory requirements for vowels which are unconstricted than for those which are produced with one or more articulatory constrictions.

Next I will analyze the C-to-V effects in more detail.

3.2.2 Effects on front vowels

3.2.2.1 Dental, alveolar and postalveolar consonants. Dental [t] causes a decrease of contact at the central palate for the vowel [i], which is consistent with the absence of contact at that same articulatory zone during the production of [t] (see Recasens, 1984). The more open vowels [e] and [ɛ] show an increase of contact at the peripheral rows when adjacent to the dental stop; this contact increase should be associated with the apicodental articulation (with simultaneous laminoalveolar contact) of the consonant as opposed to the lowered position of the blade and predorsum for the two vowels.

Alveolar [s] and postalveolar [ʃ] cause similar effects on [i]. On the one hand, rows 2 and 3 show some increase of contact; X-ray data reported by Carney and Moll (1971) also indicate that the predorso-mediodorsal constriction for [i] in the sequence [hiza] is narrower than that in the sequence [hiva]. On the other hand, the other rows (1, 4 and 5) show some decrease of contact; in an EPG study, Ohala and Kawasaki (1979) also found that [s] causes some loss of contact at the central rows of the artificial palate during the production of the vowel [i]. These C-to-V effects in linguopalatal contact suggest that the two fricative consonants are subject to the following articulatory requirements: (a) achievement of a laminal constriction, as suggested by the increase in contact on rows 2 and 3 during the vowel; (b) concave shaping of the medial groove, as suggested by the decrease in contact on rows 4 and 5 during the vowel. Indeed, contrary to traditional descriptions, there is good evidence for the primary place of constriction for [ʃ] being rather lamino-postalveolar than palato-alveolar (Hála, 1962; Koneczna and Zawadowski, 1951; Recasens, 1986), and for alveolar [s] being rather laminal than apical (for British English, Bladon & Nolan, 1977; for Catalan, Recasens, 1986). The cross-contextual stability of the medial groove configuration for [s] and, more so, [ʃ] is also documented (for American English, Carney, & Moll, 1971; for Catalan, Recasens, 1986).

When adjacent to the more open front vowels [e] and [ɛ], coarticulatory effects associated with the formation of the primary constriction (for both consonants) and with the shaping of the medial groove (for [s]) are still found. In addition, [ʃ] causes some increase of dorsal contact at the central rows since this consonant is produced with some tongue dorsum raising, and the tongue dorsum position for [e] and [ɛ] is now lower than that for [i]. On the other hand, [s] causes some positive coarticulatory effects at the periphery (row 1) since the tongue front is higher for this consonant than for the vowels [e] and [ɛ].

3.2.2.2 Velarized alveolar consonant. The consonant [ɫ] in Catalan is produced with a double articulation, namely, a primary apicoalveolar closure and a secondary dorsal constriction at the velar region or at the upper pharynx. A large cavity is found between these two places of articulation since the tongue dorsum is actively lowered at that vocal tract location (see Giles & Moll, 1975, and Fant, 1960, for similar realizations of velarized [ɫ] in American English and in Russian, respectively).

When adjacent to front vowels, this consonant has a negative effect on all rows. Notice moreover that coarticulation is larger at the central rows than at the periphery of the palatal surface. Ohala and Kawasaki (1979) also found a decrease of central contact for [i] when adjacent to American English [ɫ]. Overall, these data reflect the existence of antagonistic tongue dorsum mechanisms during the articulation of front vowels and velarized [ɫ]: while front vowels require fronting and raising of the tongue dorsum, velarized [ɫ] requires tongue dorsum lowering to bring about a "dark" acoustic quality and lowering of the sides of the tongue to allow lateral airflow.

3.2.2.3 Alveolopalatal, palatal and postpalatal consonants. These three consonantal categories are all produced with dorsal contact in the centropalatal zone. Their precise place of articulation is as follows (see Recasens, 1984, 1986 for the same speaker): postalveolo-prepalatal with the predorsum in the case of alveolopalatal [ɲ]; prepalato-mediopalatal with the predorsum and mediodorsum in the case of palatal [j]; postpalatal with the postdorsum in the case of postpalatal [k] (i.e., fronted [k] with adjacent front vowels).

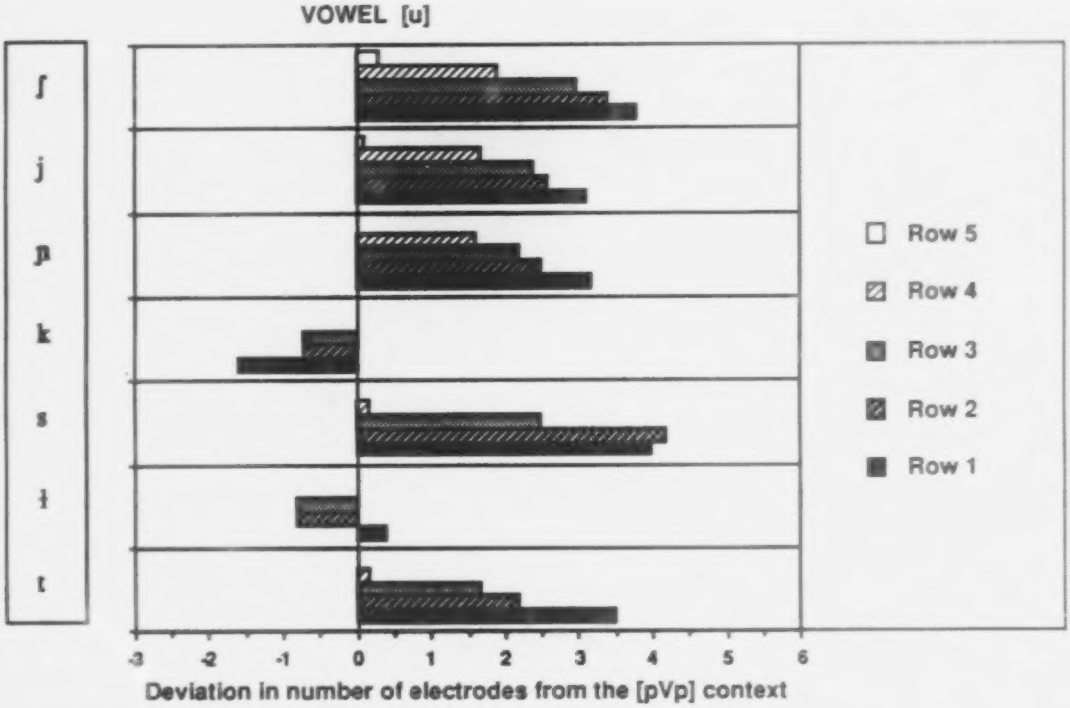


Figure 17. **Amount of deviation in number of electrodes for [u] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

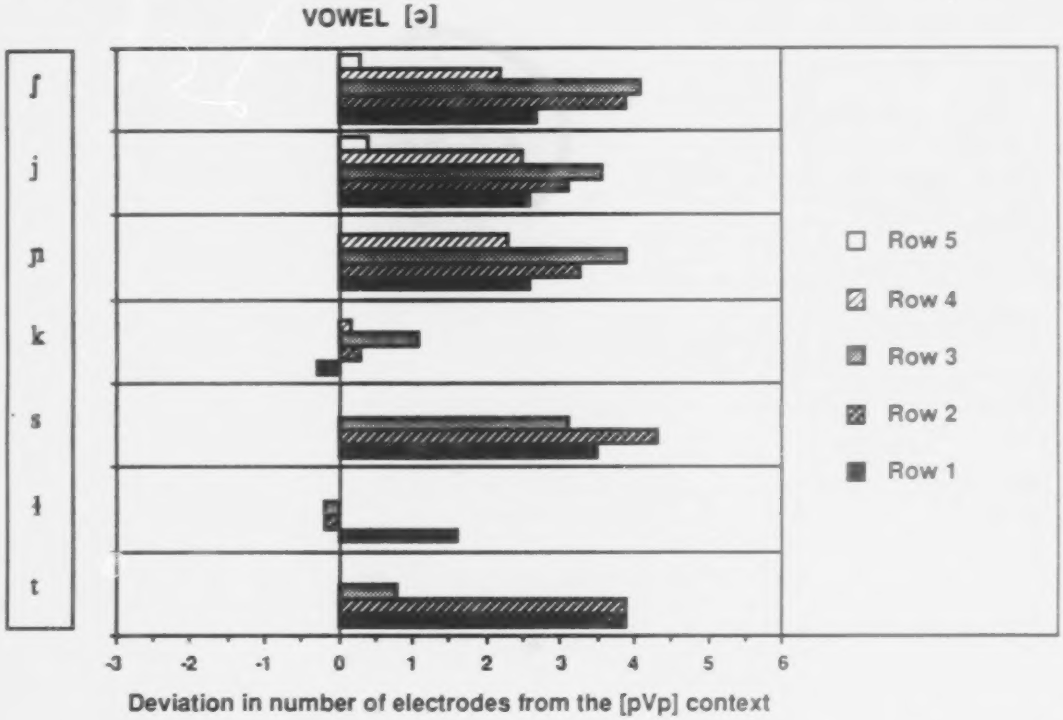


Figure 18. **Amount of deviation in number of electrodes for [ə] with respect to the neutral context [pVp] as a function of row number and consonantal context.**

When adjacent to [i] the three consonants cause a decrease of contact at the peripheral rows, more so for [j] and [k] than for [ɲ]. This effect is consistent with the fact that the two former consonants are produced with less dorsal fronting than the latter. Moreover, [ɲ] (but not [j] nor [k]) causes some increase of contact at the centropalatal zone. This finding is not in agreement with cross-consonantal differences in place of articulation; thus, negative effects ought to be larger for [ɲ] than for [j] and front [k] if lingual pressure is higher at the front palate for [ɲ], and at the mediopalate and postpalate for [j] and front [k]. Instead, these data suggest that articulatory control for [ɲ] is exerted upon the entire tongue dorsum and is not directed towards the front dorsum only.

Coarticulatory patterns on [e] and [ɛ] are similar to those on [i]. Overall, effects at the peripheral rows are mainly negative while those at the central rows are mainly positive. This is the expected trend since all these consonants share a dorsal constriction at the central palate. Positive effects at the center of the palate tend to increase as the vowel becomes lower (see § 3.2.1 for a rationale). As for [i], negative effects at the peripheral rows for these two vowels are less considerable with adjacent [ɲ] than with adjacent [j] and front [k], while positive effects at the central rows show the opposite trend.

3.2.3 Effects on back vowels and the schwa

3.2.3.1 Dental, alveolar and postalveolar consonants. For back vowels, effects at the peripheral rows are positive and quite large, and decrease in the progression [s]>[ʃ]>[t]. Thus the front dorsum is fronted and/or raised during the production of back vowels when adjacent to consonants involving a lingual closure or constriction in the front regions of the vocal tract. Consonantal effects at the center of the palate increase as the vowel becomes higher and, as expected, are larger when the adjacent consonant is [ʃ] than when it is [s] or [t].

Positive effects on [a] reflect coarticulation in tongue dorsum raising. Carney and Moll (1971) report X-ray data showing more tongue-dorsum raising for [a] when adjacent to [s] than to [v] (sequences [hisa] and [hiva]); Gay (1974) reports X-ray data showing similar effects in the progression [t]>[k]>[p] (sequences [VCa]).

There is also evidence for coarticulation in lingual fronting and raising upon back rounded vowels in these consonantal contexts. Stone et al. (1987) report ultrasound data showing a similar tongue dorsum configuration for [i] and [u] in *dVd* sequences for some speakers of American English; according to X-ray data in Perkell (1969), the vertical distance between the dorsum of the tongue and the mediopalate in the context [hɔtV] is higher for [u] than for [ɪ] and [ɛ]; X-ray data reported by Carney and Moll (1971) show more fronting for [u] in the sequence [husi] than in the sequence [huʃi]; Butcher and Weiher (1976) give EPG data showing lateral contact up to the front alveolar zone (one speaker) and up to the front prepalatal zone (another speaker) in the case of [u] in the sequence [utu]; according to X-ray data reported by Kiritani et al. (1977), the tongue stretches out and becomes flat during the production of [o] and [u] in symmetrical CVC sequences with dentoalveolar consonants.

The chances that the front dorsum configuration for back vowels is affected by the adjacent consonants are high in CVC symmetrical environments, as shown by myself in this paper, and by Stone et al. (1987) and Kiritani et al. (1977). Indeed data collected in other contextual conditions may reveal no consonantal effects. Thus, according to X-ray data from the literature, [u] may be articulated with a similar tongue dorsum configuration at a dorsovelar place of constriction when adjacent to consonants differing in degree of fronting ([udV] and [ugV] (Öhman, 1966); [utV], [upV] and [ukV] (Gay, 1974)).

A large degree of variability in linguopalatal contact for [u] is due to lingual fronting effects caused by adjacent consonants articulated at the front of the mouth (see Figure 9). This phenomenon may be associated with lip rounding, which is more considerable for this vowel than for the other back rounded vowels. It can be hypothesized that speakers do not care much about the precise location of the lingual constriction because the spectral properties of [u] are already accounted for by the lip rounding gesture. If so, it seems a natural strategy to allow for C-to-V coarticulation in tongue positioning along the palatal surface.

Larger effects in front dorsum activity when the adjacent consonant is a fricative than when it is a stop may result from the high requirements on the shaping of the tongue body in the case of [s] and [ʃ]. They might also be associated with more anticipatory lip rounding during [s] and [ʃ] than during [t].

Effects on the schwa are similar to those on [u] but not identical, because neither is making strong demands on the tongue. Positive deviations are most peripheral for [t] and most central for [ʃ],

[s] falling in between. These trends reflect presumably the contribution of the blade in the production of [s] and [ʃ], of the dorsum in the production of [ʃ], and of the apex in the production of [t]. Overall effects on [ə] reflect quite well the articulatory properties of the adjacent consonant.

3.2.3.2 Velarized alveolar consonant. This consonant causes few coarticulatory effects on the pharyngeal vowels [a], [ɔ] and [o]. When adjacent to [u] and the schwa, it has some negative effect on rows 2 and 3 (more so when V=[u] than when V=[ə]) and some positive effect on row 1 (more so when V=[ə] than when V=[u]).

The absence of coarticulatory effects on pharyngeal vowels is presumably due to the similarity in tongue dorsum configuration between those vowels and velarized [ɫ]; indeed, all these articulations are produced with a back dorsal constriction and a lowered front dorsum. Larger effects on the central rows for [u] than for the schwa are associated with more tongue dorsum raising in the case of [u]; larger effects on the peripheral rows for the schwa than for [u] indicate that the schwa may be more neutral to tongue fronting induced by [ɫ] (see also coarticulatory effects with adjacent [t]).

3.2.3.3 Alveolopalatal and palatal consonants. In this contextual condition, positive coarticulatory effects increase as the vowel becomes higher, mainly at the center of the palate (see § 3.2.1 for a rationale). In all cases except for [a] (which, like [ɔ], shows very small effects) coarticulation is larger at the peripheral rows than at the central rows. This trend suggests that, similarly to adjacent dentoalveolars, adjacent palatal consonants cause some stretching of the front dorsum during back rounded vowels and pull it forward within the oral cavity.

In general, effects on the schwa are as large as on [u]; for that vowel, however, they decrease in the progression 3rd row> 2nd row> 1st row> 4th row> 5th row. Notice also that effects on rows 2 and 3 are higher for the alveolopalatal [ɲ] than for the palatal [j] while those on rows 4 and 5 are larger for the palatal than for the alveolopalatal. In summary, differently from front vowels (see § 3.2.2.3), effects on [ə] reflect quite well the place of constriction of the adjacent consonants, which is more fronted and less central for alveolopalatals than for palatals.

3.2.3.4 Velar consonant. This consonant causes slight negative effects on row 1 in the case of back vowels and the schwa, and additional effects on rows 2 and 3 in the case of [u] (negative) and the schwa (positive). The presence of negative effects indicate that [k] with back vowels is produced with a velar place of articulation; thus a very retracted tongue position for the consonant causes some retraction during the vowel when compared to its neutral configuration in the *p*V*p* environment.

4. ARTICULATORY-ACOUSTIC RELATIONS

As pointed out in the Introduction, another goal of this study is to establish a correlation between the EPG data and the acoustic data.

Figure 19 shows F2 data for all vowels in all consonantal contexts. Differences between these F2 values and those in the *p*V*p* context are not given since F2 for a vowel preceded and followed by p reflects other articulatory factors besides tongue activity. It should also be said that the correlation between the articulatory and acoustic measurements will be necessarily low for those vowels showing little palatal contact since the lingual configuration cannot be inferred with precision from the EPG data in these circumstances.

4.1 Front vowels

A systematic relationship between the F2 data and the EPG data obtains in the case of front vowels with adjacent velarized [ɫ]. In this case, negative C-to-V effects in degree of dorsopalatal contact (see § 3.2.2.2) cause a significant decrease in F2. Wood (1982) has also shown that large openings at the place of the palatal constriction cause dramatic F2 changes while location perturbations along the prepalate and mediopalate do not cause appreciable F2 variations.

There may also be a direct relationship between positive C-to-V effects in dorsopalatal contact (EPG data; see § 3.2.2.3) and a high F2 frequency in the case of [e] and [ɛ] when adjacent to alveolopalatal [ɲ], palatal [j] and postpalatal [k]. Of course it could be argued instead that dentoalveolars and labials (but not alveolopalatals and palatals) cause some deviation (i.e., lowering) from the ideal F2 values for those target vowels (Stevens & House, 1963).

Notice that the fricatives [s] or [ʃ] tend to cause a lower F2 frequency than the stop of a similar place of articulation (i.e., [t]) at the midpoint of [i] and [e] (see also Stevens and House, 1963). A rationale for this finding is that fricatives cause some tongue dorsum lowering along the median line during the production of the vowel and thus contribute to an F2 decrease. Some evidence for such effects has been reported in § 3.2.2.1.

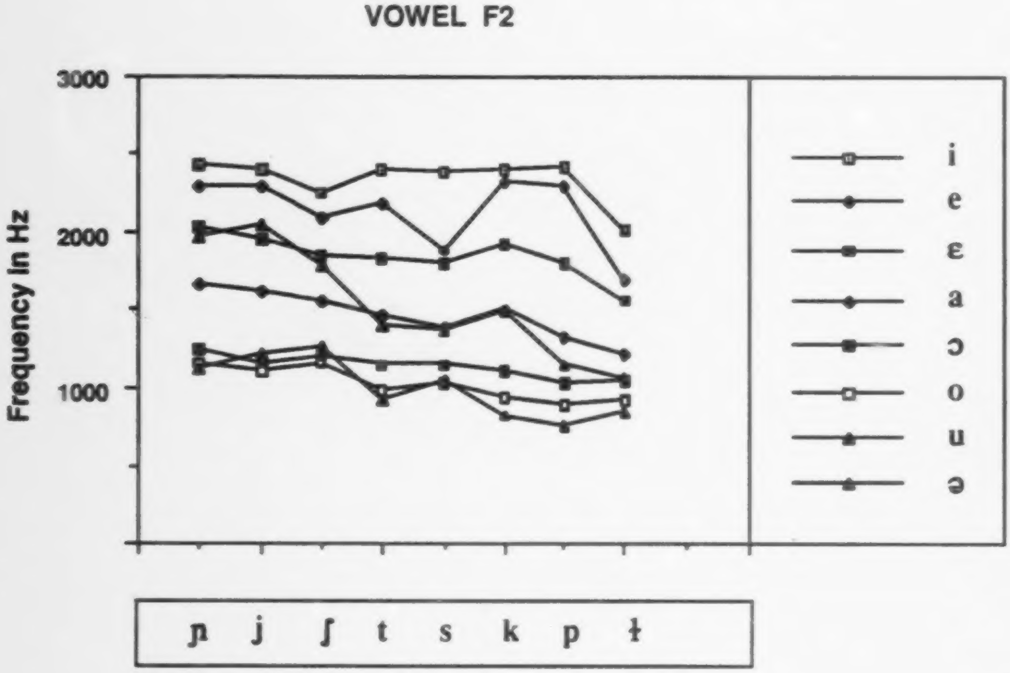


Figure 19. F2 data for all vowels in all symmetrical CVC contexts.

4.2 Back vowels and the schwa

Among back vowels, positive C-to-V effects in the EPG data have some clear acoustic consequences (also Stevens & House, 1963). Thus, vocalic F2 is highest when the adjacent consonant involves a palatal constriction, less high when it is articulated in the front of the mouth but shows no contact at the central palate, and lowest when it is a velar or the velarized [ɫ]. Consistently with the data reported here, Wood (1982) has shown that advancing the dorsal constriction for [u] from the back or mid soft palate to the front soft palate or more forward causes a great F2 increase; Maeda (in press) reports a similar finding for the vowel [o]. Notice that this consonant-dependent F2 increase could also reflect lip unrounding during the production of back rounded vowels.

According to Figure 19, fricatives (i.e., [s]) tend to cause a higher F2 of back vowels than stops of similar articulatory characteristics (i.e., [t]) (also Stevens & House, 1963, for American English).

F2 effects for the schwa are very similar to those observed on the palatal surface. They can be compared to those for [a] and show a larger spread of F2 frequencies across consonantal contexts.

5. SUMMARY AND CONCLUSIONS

Patterns of linguopalatal contact across consonantal contexts show differences in contact size between vowels of different height within the front series ([i], [e], [ε]) but not within the back series ([u]> [a]> [o], [ɔ]). This apparent anomaly was attributed to the position of the front dorsum, which is higher for low back vowels than for higher back vowels. A systematic contrast in constriction location on the palatal surface was obtained for front vs. back vowels which is consistent with the tongue regions under active control in the two groups. This grouping of vowels into a front and a back category has been shown to be not only related to patterns of linguopalatal contact but also to patterns of articulatory variability. Variability for [ə] is the highest of all vowels.

The finding that [i] is subject to a low degree of variability in palatal contact is consistent with the assumption of [i] being a quantal vowel (Perkell & Nelson, 1985). According to this assumption, tongue front activity shows little variability along the vertical dimension because the sides of the tongue blade are being pushed against and restrained by the hard palate; on the other hand, horizontal variability at the front and, more so, at the back of the tongue dorsum is caused presumably by contraction of the posterior genioglossus. Moreover, findings reported here suggest that coarticulatory resistance on the entire tongue dorsum is also found when the tongue region under control is the predorsum and/or mediodorsum instead of the blade.

High cross-consonantal SD values in EPG activity for [u] and [ə] suggest a clear trend for the degree of coarticulation to vary inversely with the degree of constraint on tongue dorsum activity. Low SD values for [a], [o] and [ɔ] when compared to those for [u] suggest that lip rounding may play a role in the presence of C-to-V effects.

As shown below, coarticulatory data reported in this paper may be useful to gain some understanding about the nature of the articulatory strategies used by speakers for the production of adjacent vowels and consonants. It should be pointed out in this respect that C-to-V effects are subject to similar production strategies to those observed in V-to-C effects. Analogously to consonants, those lingual regions which are not involved in the formation of a vocalic constriction are left freer to coarticulate than those which are actively involved.

Effects on front vowels are quite small due to the requirements on the tongue dorsum to achieve a palatal constriction. Overall, tongue dorsum raising effects increase as the vowel becomes lower, in line with differences in articulatory constraint. The most significant consonantal effects are those caused by velarized [ɫ], stop vs. fricative dentoalveolars, and alveolopalatals vs. (post)palatals. In general, those articulatory contrasts associated with palatal constriction width have an effect on F2 frequency.

The number and size of the positive coarticulatory effects are larger for back vowels and [ə] than for front vowels; moreover, differently from front vowels, effects among back vowels increase with vowel height at the periphery and at the center of the palate. Large effects on [ə] are found all over the palatal surface, including the central palate. Back vowels and the schwa show large positive coarticulatory effects which reflect raising (typically for [a]) and fronting (typically for rounded vowels) of the tongue dorsum as a function of adjacent dentoalveolar and palatal consonants; [k] and velarized [ɫ] however cause negative or no positive effects. These articulatory differences have an effect on F2 frequency.

Overall, these effects reflect contrasts in degree of articulatory constraint associated with different phonetic segments. Large effects on back vowels from dentoalveolar and palatal consonants show that the constriction requirements on the front dorsum (for the consonant) dominate in the absence of articulatory activity (for the vowel); on the other hand, coarticulatory effects from velarized [ɫ] on back vowels are small since the front dorsum configurations for that consonant and those vowels are highly similar. In the case of front vowels, high requirements on tongue dorsum lowering for velarized [ɫ] override to some extent tongue dorsum raising for front vowels; also, coarticulatory effects from fricatives on front vowels are very precise because fricative production demands high requirements on the tongue configuration.

REFERENCES

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech*, *25*, 151-173.

Baer, T., Alfonso, P. J., & Honda, K. (1988). Electromyography of tongue muscles during vowels in /əpVp/ environment. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, University of Tokyo, *22*, 7-19.

Bladon, R. A. W., & Nolan, F. J. (1977). A videofluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, *5*, 185-193.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of tip and blade alveolars in English. *Journal of Phonetics*, *4*, 59-74.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, *23*, 193-202.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press.

Gay, T. (1974). A cinefluorographic study of vowel production. *Journal of Phonetics*, *2*, 255-266.

Giles, S. B., & Moll, K. L. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*, 206-227.

Hála, B. (1962). Uvedení do fonetiky češtiny na Obecně fonetickém základě. Prague.

Hamlet, S. L., Bunnell, H. T., & Struntz, B. (1986). Articulatory asymmetries. *Journal of the Acoustical Society of America*, *79*, 1164-1169.

Harshman, R., Ladefoged, P. & Goldstein L. (1977) Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, *62*, 693-707.

Kiritani, S., Itoh, K., Hirose, H., & Sawashima, M. (1977). Coordination of the consonant and vowel articulations—X-ray microbeam study on Japanese and English. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, University of Tokyo, *11*, 11-21.

Koneczna, H., & Zawadowski, W. (1951). *Przekroje Rentgenograficzne glosek Polskich*. Warsaw.

Maeda, S. (in press). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling*.

Marchal, A., & Espesser, R. (1989). L'asymétrie des appuis linguo-palatins. *Journal d'Acoustique, 2*, 53-57.

Ohala, J., & Kawasaki, H. (1979). The articulation of [i]'s. Paper presented at the 98th Meeting of the Acoustical Society of America, Salt Lake City.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151-168.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge: The MIT Press.

Perkell, J. S., & Nelson, W. L. (1985). Variability in production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America, 77*, 1889-1895.

Pols, L. C. W. (1977). *Spectral analysis and identification of Dutch vowels in monosyllabic words*. Soesterberg: Institute for Perception TNO.

Recasens, D. (1984). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics, 12*, 61-73.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech, 28*, 97-114.

Recasens, D. (1986). *Estudis de fonètica experimental del català oriental central*. Barcelona.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulation by consonantal context: An acoustical study. *Journal of Speech and Hearing Research, 6*, 111-128.

Stone, M., Morrish, K. A., Sonies, B. C. & Shawker, T. H. (1987). Tongue curvature: A model of shape during vowel production. *Folia Phoniatrica, 39*, 302-315.

Wood, S. (1982). Radiographic and model studies of the palatal vowels. X-ray model studies of vowel articulation. *Working Papers in Linguistics, 23*, 119-155, Lund University.

FOOTNOTES

[*] *Journal of Phonetics*, in press.

[†] Also Universitat Autònoma de Barcelona, Department de Filologia Catalana, Bellaterra, Barcelona.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 149-176

Coarticulation Is Largely Planned*

D. H. Whalen

Is coarticulation a reflection of planning an utterance or the automatic effect of producing speech? This question was examined by requiring speakers to begin reading nonsense aCV or əbVCa strings (where C was /b/ or /p/ and V was /i/ or /u/) aloud before seeing the entire utterance. Those segments known before articulation began exerted normal anticipatory coarticulatory influences, while those seen after utterance onset usually did not. Perseverative coarticulation was, for the most part, present in both conditions, showing either that it is indeed a result of the mechanical nature of the speech organs or that the amount of planning necessary can be carried out in the time used for the first syllables. There was one anticipatory effect which began to emerge (even though unplanned) just before the stop closure, supporting an articulatory overlap explanation for some of the coarticulation. From the greater magnitude and extent of the coarticulatory influence of segments known in advance, we can infer that, for the most part, coarticulation is planned.

INTRODUCTION

Coarticulation, the influence of one phoneme on another, has received a great deal of attention in the phonetics literature (e.g., Fowler, 1980, 1984; Keating, 1985; Kent & Minifie, 1977; Liberman & Mattingly, 1985; Liberman & Studdert-Kennedy, 1978). This attention is well deserved, since the widespread and complex set of influences among phonemes results in the lack of simple acoustic invariance in the speech signal which has made the field of phonetics so challenging. But while coarticulation clearly influences the output of the speech system, does it do so as part of the motor program or as a consequence of executing that program? The work of Sternberg and his colleagues (Monsell, 1986; Sternberg, Monsell, & Monsell, 1980) has provided evidence that the motor program for speech is encapsulated so that the complexity of the program (at least within a stress foot) does not affect its initiation time in speeded productions. While the existence of these programs seems well-established by the experiments done to date, the details of the program are only partially evident. In particular, it is unclear whether the program consists of a sequence of fully-specified phoneme-sized commands, or instead a series of gestures which can be executed with more simultaneity. If the former is the case, then the effects of coarticulation will appear to be the output of explicit contextualization rules. If the latter, then coarticulation could be seen as a direct consequence of the output of speech, where activations for the various gestures produces the contextualization by means of an overlapping of their realizations.

The question of planning has taken on a slightly different guise in the phonetic literature, namely, whether the modifications of coarticulation are due to feature-spreading (as in Henke, 1966, and Keating, 1985, 1988) or by temporal concurrence, such as "shingling" of phones (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967:441), overlap of prominence curves for phones (Fowler & Smith, 1986) or the phasing of gestures (in the Articulatory Phonology of Browman & Goldstein, 1985, 1986). The feature-spreading approach inherently postulates that coarticulatory effects will be equally large throughout the production of the affected phoneme (cf. Manuel 1987:18-19).

This research was supported by NSF Grant BNS-8111470 and NICHD Grant HD-01994 to Haskins Laboratories. Invaluable assistance in data collection and analysis was provided by Diana Matson, Mark Pitt, and Yi Xu. Comments on earlier drafts by Caroline Smith, Carol A. Fowler, Vincent L. Gracco, Sharon Y. Manuel, Daniel Recasens, Elliot L. Saltzman, Anders Löfqvist, Alvin M. Liberman, Catherine P. Browman, Louis M. Goldstein, Margaret Hall Dunn, Rena A. Krakow, Suzanne Boyce, Kevin G. Munhall, Patrice Speeter Beddor, Mary Beckman and an anonymous reviewer were most helpful.

However, both Henke and Keating have introduced interpolating mechanisms which allow for an increase in the effects as the influencing phoneme is approached. In addition, Keating proposes that some phones will be unspecified for a feature, allowing for continuous change in its production. Features are claimed to spread as far back as they can, before reaching a contradictory specification. Conversely, the overlap models inherently predict severe limits on the duration over which coarticulation can occur. Fowler (1981:38), for example, asserts that vowels should affect other stressed vowels only at the onset or offset (though certain components, such as lip rounding, might not be so limited). Measurements by Magen (1989), however, show vowel-to-vowel coarticulation throughout both stressed and weakly stressed vowels, even though the affecting vowel was separated from the affected one by a syllable containing the vowel schwa. Thus some modification of overlap models is already called for, and the present experiments will add to the evidence needed to make those modifications.

A special case of overlap is the frame model of Bell-Berti and Harris (1981, 1982). In this model, coarticulatory overlap is dependent on such properties of speech as stress and rate of articulation, but is independent of the number and composition of preceding segments. Thus lip-rounding activity will precede vowel onset by the same amount of time whether it is preceded by two or many segments, and whether these segments are intrinsically long or short. In their model, then, what is planned is not coarticulation per se but simply the onset of the upcoming vowel, with the motor commands for the gesture affecting the signal more or less depending on the competing signals of the preceding gestures. Similarly, Articulatory Phonology postulates that the shape of gestures is insensitive to context, and that contextual variation is the result of the blending of the various levels of activation. In this sense, even though coarticulation can be seen in the signal, it has not been explicitly planned.

In all these theories, the input to the system is assumed to be complete, that is, all the phonemes of the utterance are known to the speaker. Yet if a speaker were to begin speaking before the whole utterance was known, the feature-spreading and overlap models make different predictions. By feature-spreading, if one phoneme is started before its context is known, then there should be no features to spread, thus coarticulatory influences should not appear. Even at the end of the production of an initial phoneme, there should be no influence, since the interpolation of the feature-spreading models is still based on features, and the initial segment does not have a feature value to interpolate from. In overlap models, if the phoneme is started before the context is known, the overlap should still appear since the dynamic curves ARE the definition of the phonemes. Even in these theories, however, there is a mechanism by which the overlap would not appear: Some articulatory gestures can be "hidden" (Browman & Goldstein, in press), that is, the articulation can be present yet there is no acoustic evidence of it. In the best case, though, the two types of theories could be distinguished on the basis of productions which were begun on the basis of incomplete information.

The present experiments looked at coarticulatory effects in simple bisyllables (abí, abú, apí, apú) and trisyllables (əbíba, əbúba, əbípa, əbúpa) which were initiated before part of the utterance was known. This was accomplished by displaying on a computer screen part of the utterance (e.g., A_U) and then filling in the missing part as soon as phonation began. In this way, various coarticulatory effects could be examined in conditions where they could be planned versus those where any effect would have to be made in mid-production. The conditions differed by the segment that was missing in the initial display. This could be either the consonant (B or P) or the final or middle vowel (I or U). Four coarticulatory effects were measured. The one associated with the vowel was the effect of shifting F2 slightly toward the F2 of an upcoming vowel (Öhman, 1966). Three effects were associated with the voicing of the consonant. The offset frequency of a vowel's F1 before a voiced stop is lower than it is before a voiceless (Lisker, 1986; Wolf, 1978). Vowel duration is longer and closure duration shorter for a voiced stop relative to a voiceless one, with the two parts of the effect usually cancelling each other out so that overall duration is unaffected (Sharf, 1962). And, in the one effect which is perseverative for both the bi- and trisyllabic utterances, the F0 after release of a voiceless stop is high relative to a voiced stop (e.g. Hombert, Ohala, & Ewan, 1979). In all these cases, it is possible to compare the two types of uncertainty, i.e., whether the vowel or the consonant was unknown at utterance onset, to see whether the coarticulatory effects emerge or not.

1.0 EXPERIMENT 1

The first experiment examined the four coarticulatory effects in nonsense bisyllabic utterances.

1.1 Method

1.1.1 Subjects

The subjects were the author (DHW), a male colleague (LMG), and a female colleague (DMM). All were native speakers of American English.

1.1.2 Stimuli

Each utterance consisted of a bisyllabic nonsense word. The first syllable was the low-central vowel /a/. The second syllable began with either /b/ or /p/ and ended with either /i/ or /u/. Stress was to be placed on the second syllable. Since the subjects were all familiar with the International Phonetic Alphabet, these were represented on the computer screen as ABI, ABU, API and APU.

1.1.3 Equipment and Procedure

Subjects produced the utterances in response to a visual display on a computer screen (Atari 800). In the first condition, only the two vowels appeared on the screen initially, with a blank between them (e.g., "A__U"). When this appeared, the subject could prepare for as long as he or she wished before initiating a response. Once the subject began phonating the /a/, a Schmidt trigger connected to the microphone output sent a signal to the computer. Once the signal was received, the consonant immediately appeared in the previously blank position, and the entire letter string remained on the screen for 1 second. The subject was instructed to incorporate that consonant into the ongoing production as rapidly and as smoothly as possible. In the second condition, the first two letters appeared on the screen (e.g., "AB__"). Here again, the subject's task was to prepare to utter the bisyllable, and then begin. As soon as he or she did, the missing vowel appeared. That vowel was then incorporated as quickly and as naturally as possible into the utterance.

In each condition, twenty repetitions of each utterance type (e.g., /abi/) were randomized. If the subject made a mistake on a particular token, there was no repetition of that item. Instead, that token was excluded from the analysis. Errors were judged by the author and a colleague independently. Subject responses were recorded on audio tape on an Otari MX-5050 reel-to-reel recorder at 7 1/2 ips. They were then low-pass filtered at 5 kHz and digitized through the Haskins PCM system at 10 kHz with 12 bit resolution for computer analysis.

All duration measurements were made interactively with the wave-form editing program WENDY. All formant measurements were based on formants extracted from the LPC parameters generated by the ILS software package. These were further corrected by hand for those instances where the program had missed or added a formant. Missing values were excluded from the analysis. F0 measurements were made by delimiting individual pitch periods in the waveform, again with WENDY.

1.2 Results

Error rates (which included mispronunciations of either the vowel or the consonant) for the three subjects for the consonant-unknown and vowel-unknown conditions were: 3.75% and 8.75% (DHW); 15.0% and 5.0% (LMG); and 0% for both for DMM. In the consonant-unknown condition, DHW and DMM lost an additional 5.0% and 3.75% of the data due to technical malfunctions. Across the two subjects who made errors, about half (12 of 26) were on the item that was missing, while the others were on the consonant or variable vowel which had been visible from the beginning. Thus the procedure as a whole, not the nature of the unknown segment, seems to be responsible for the pattern of errors.

The results for the vowel-to-vowel effect on F2 are shown in Figure 1. The top panel represents the data for the consonant-unknown condition, collapsed across subject and consonant. The middle panel is that of the vowel-unknown condition, while the bottom panel superimposes the two. The first two data points for each function in each panel represent the first two LPC analysis frames (10 ms apart) at the beginning of the vowel. The next six points represent the frames just before the closure for the stop. In a few cases where the vowel was short, the same data appeared in the points before closure and the initial two.

In the consonant-unknown condition, the upcoming vowel is known. And, as can be seen in the separation of the functions in the top panel, that vowel exerts an influence on the F2 of /a/, in much the same way as originally found by Öhman (1966). By contrast, when the vowel is not known (middle panel), it affects neither the beginning nor the last vestiges of the /a/ before closure (i.e., the functions are hardly separable). Separate analyses of variance were performed on the beginning frames and the final frames in each of the two conditions, with the factors of Speaker, Following Vowel, and Token (which allowed for missing values). The effect of Following Vowel was significant in the consonant-unknown condition for both the beginning frames ($F(1,205) = 8.12$, $p < .01$) and the final frames ($F(1,205) = 4.35$, $p < .05$).

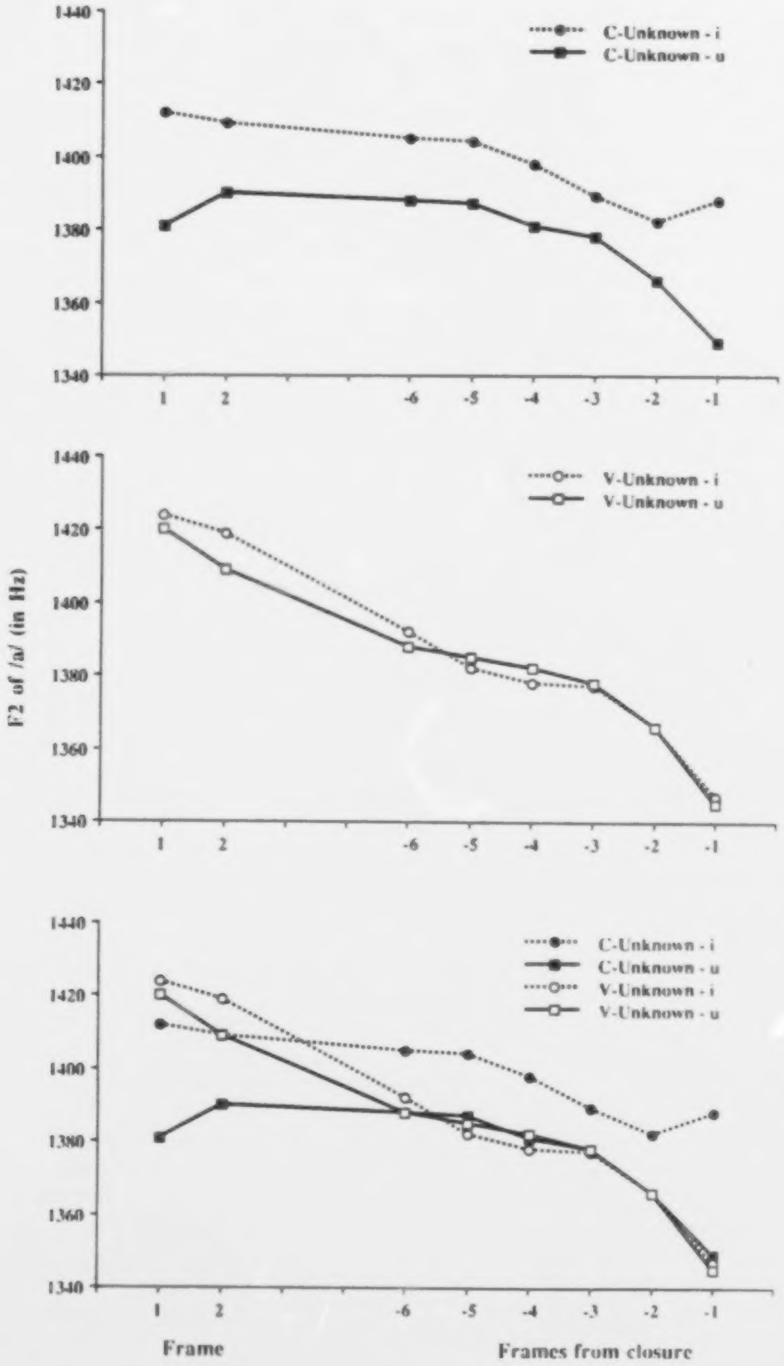


Figure 1. Effect of an upcoming vowel on the F2 of initial /a/, Experiment 1. In the consonant-unknown condition (top panel), the vowel is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is not. The lower panel superimposes the upper two panels. The first two points are the analysis frames (10 ms per frame) at the beginning of the vowel. The last points are the frames just before closure.

In the vowel-unknown condition, neither location showed a significant difference ($F(1,217) < 1.0$, n.s., for both). While speaker was a significant main effect in each analysis, indicating that the three subjects had different vowel spaces and/or vocal tract shapes, it entered into only one significant interaction. In the first frames of the consonant-unknown condition, the speaker by vowel interaction was significant ($F(1,205) = 5.45$, $p < .01$). While DHW and LMG showed the pattern evident in the first two points of Figure 1, DMM had a non-significant difference in the opposite direction. This is at odds with the fact that she behaved like the other two subjects later in the vowel. Thus the effect of upcoming vowel on F2 of /a/ was clearly present in the consonant-unknown condition (for two subjects) and completely absent from the vowel-unknown condition.

The effect of consonant voicing on vowel duration was equally straightforward (Figure 2). In the vowel-unknown condition, where the consonant was known, the usual lengthening of the vowel and shortening of the closure for the voiced stop relative to the voiceless was found. In the consonant-unknown condition, by contrast, there was no difference between the vowel duration (including the voiced transitions) before the two stops. In addition, there was no difference in the closure durations. This was in spite of the fact that the vowel duration was nearly twice as long as it was in the vowel-unknown condition. Separate analyses of variance were performed on vowel duration, closure duration, and total duration (vowel plus closure) for each of the two conditions, with the factors of Subject, Voicing and Token. Total duration did not differ with the voicing of the consonant in either condition ($F(1,205) < 1.0$ and $F(1,217) = 2.47$, n.s., for the consonant-unknown and vowel-unknown conditions respectively). Both vowel duration ($F(1,217) = 70.33$, $p < .001$) and closure duration ($F(1,217) = 41.27$, $p < .001$) differed in the vowel-unknown condition, but neither vowel duration nor closure duration differed in the consonant-unknown ($F(1,205) = 1.97$ and 1.64, n.s.). Speaker was a highly significant factor, due in large part to the extreme durations in DMM's productions. Her closure durations were more like those of geminate stops than of typical American English stops. Nonetheless, none of the interactions with the voicing factor with speaker were significant, indicating that even DMM's extremely conservative production style still allowed the duration effect to emerge. The effect of the stop's voicing on the F1 of the preceding vowel was strong when the consonant was known and virtually absent otherwise.

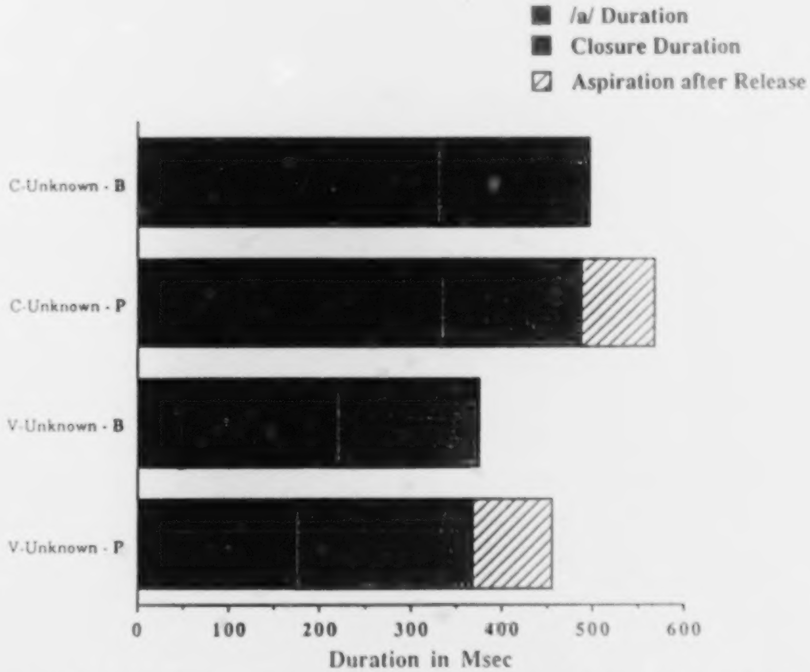


Figure 2. **Effect the voicing of an upcoming consonant on the duration of initial /a/ and the stop closure, Experiment 1. In the consonant-unknown condition (top two bars), the consonant is not known at utterance initiation time. In the vowel-unknown condition (lower two bars), it is.**

In the middle panel of Figure 3, the last six frames before closure for the stop show that the F1 of /a/ reaches a lower frequency before the voiced stop than before the voiceless when the consonant was known ($F(1,217) = 101.40$, $p < .001$). (This effect does not appear to be due to the different durations of the vowels. Even if we shift the F1 from the voiced stop context by the difference in duration of the /a/, there is still a sizable effect of voicing on F1.) In the top panel, for the consonant-unknown condition, there is no difference for all the points taken together ($F(1,205) = 1.30$, n.s.). The apparent divergence in the last two frames before closure in the consonant-unknown condition, however, does reach significance ($F(1,205) = 4.17$, $p < .05$ and 10.86, $p < .01$, for the next-to-last and last frames respectively). So, unlike the vowel effect on F2, the voicing effect on F1 does show up at the very last instant.

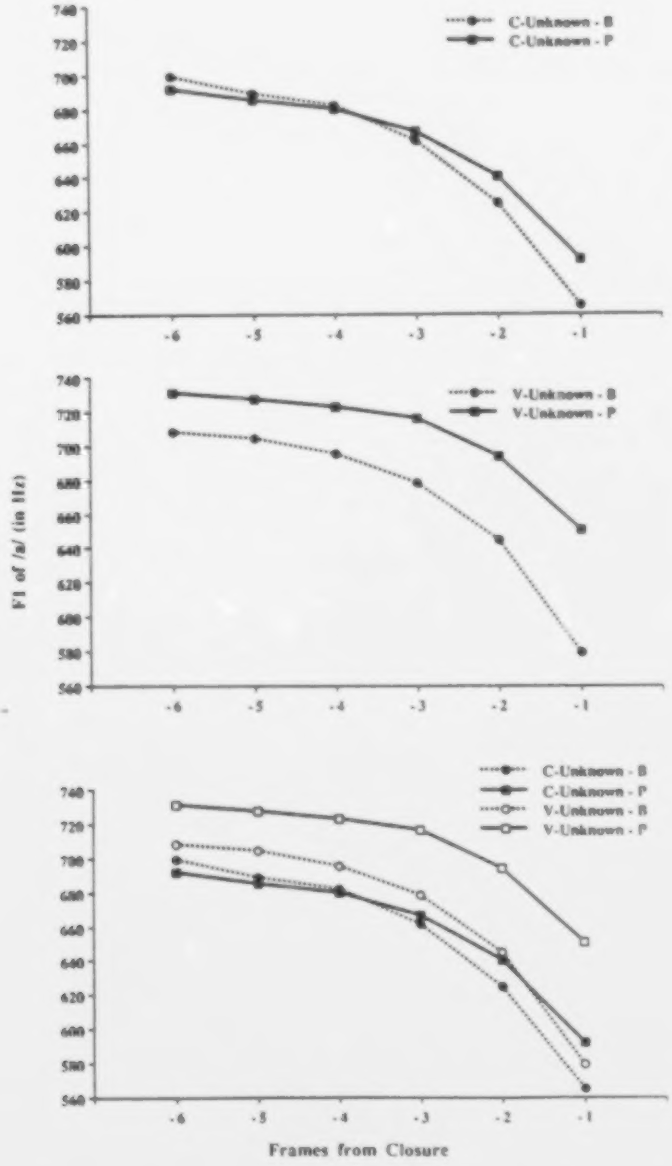


Figure 3. Effect the voicing of an upcoming consonant on the F1 of initial /a/, Experiment 1. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels.

First acoustic evidence (in the formant transitions) of the production of the missing segment appears about 300-320 ms from utterance onset in the consonant-unknown condition. The first solid evidence in the vowel-unknown condition appears at the release of the stop, around 360 ms, but the actual onset could have been up to 140 ms earlier, since the closures are, on average, that long. It is possible that the time to resolve the identity of the missing segment was the same in both cases, since both involved a choice between two phonemes.

In an earlier examination of consonant voicing on F1, Summers (1987) found that F1 was affected throughout the syllable's duration. Analysis of the F1 in the first frames of the present stimuli showed no effect of consonant voicing in either condition.

The effect of the voicing of the stop on the following vowel's F0 was measured for each pitch period in the first 50 ms, and then one pitch period in the middle of the vowel. The "mid-vowel" pitch period was the one which included the point 100 ms after the onset of voicing for the voiceless stops; for the voiced stops, it was the pitch period which included the point 100 ms past the average VOT for the voiceless stops for that subject for that condition. This point was chosen to be somewhat later than Ohde's (1984) "target" pitch period, so that any effects would have to be somewhat stronger than the ones he found. In terms of absolute duration, this "mid-vowel" pitch period occurs somewhat earlier than the midpoint, but it was always after the most intense part of the vowel. For the statistical analysis, an equal number of pitch period values from each condition was selected, with a different number for each speaker. The goal was to analyze the first few pitch periods in such a way that all the utterances would have an equal number of measurements, even though this meant leaving out some of the F0 pattern in the first 50 ms. The number of periods chosen obviously varied with the F0 itself, so that there were more data points for the female, fewer for the higher pitched male (DHW) and least for the lower pitched male.

As can be seen in Figures 4, 5 and 6, the voiceless consonant elicited a higher F0 than the voiced in both conditions. For DMM (Figure 4), there was no main effect of condition either on the first pitch periods or on the mid-vowel ($F(1,149) < 1.0$ for each), while the effect of consonant was strong ($F(1,149) = 117.71$ and 743.91, $p < .001$, for early and mid-vowel respectively). There was an interaction of condition and consonant in the mid-vowel measurement; the effect was present in both, but slightly larger in the consonant-unknown condition. For LMG (Figure 5), there was a main effect of condition for the early pitch periods ($F(1,134) = 84.94$, $p < .001$), though it did not interact with consonant. The consonant also strongly affected the initial pitch periods ($F(1,134) = 166.32$, $p < .001$). For the pitch period in the middle of the vowel, there was no effect of condition ($F(1,134) = 1.24$, n.s.), and a strong one for consonant ($F(1,134) = 61.85$, $p < .001$) and an interaction between the two ($F(1,134) = 4.56$, $p < .05$). Separate analyses of the two conditions revealed that the pitch period in the middle of the vowel differed with consonant in both conditions but with greater magnitude in the consonant-unknown condition. For DHW (Figure 6), there was an effect of condition and consonant for both the early and the midpoint pitch periods ($F(1,138) = 6.95$, $p < .01$ and $F(1,138) = 372.96$, $p < .001$ for the early, $F(1,138) = 12.22$, $p < .001$ and $F(1,138) = 478.21$, $p < .001$ for the midpoint). The interaction was significant only in the mid-vowel pitch period. Although the F0 was somewhat higher in the vowel-unknown condition, the early effect of the consonant was equally large in both conditions. In the middle of the vowel, the effect of the consonant's voicing was somewhat smaller in the consonant-unknown condition than in the vowel-unknown condition.

Thus, whether the consonant had been planned for in advance or not, the typical F0 pattern emerged both at the beginning of the vowel and well after both the aspiration and release of the consonant. Some of the difference in the mid-vowel F0 might be due to the slight difference in duration of the final vowels. The syllables with voiceless stops tended to be about 35 ms longer than those with voiced stops. This means that the measurements for the voiced syllables were somewhat further along in a declining intonation pattern. However, it does not appear that the duration differences are large enough to account for the whole F0 difference. However, the difference in the size of the effects between conditions may be artifactual rather than real: The duration difference is larger in the vowel-unknown (45 ms) than the consonant-unknown (25 ms) condition. On the whole, then, this perseverative effect does not show a sensitivity to planning.

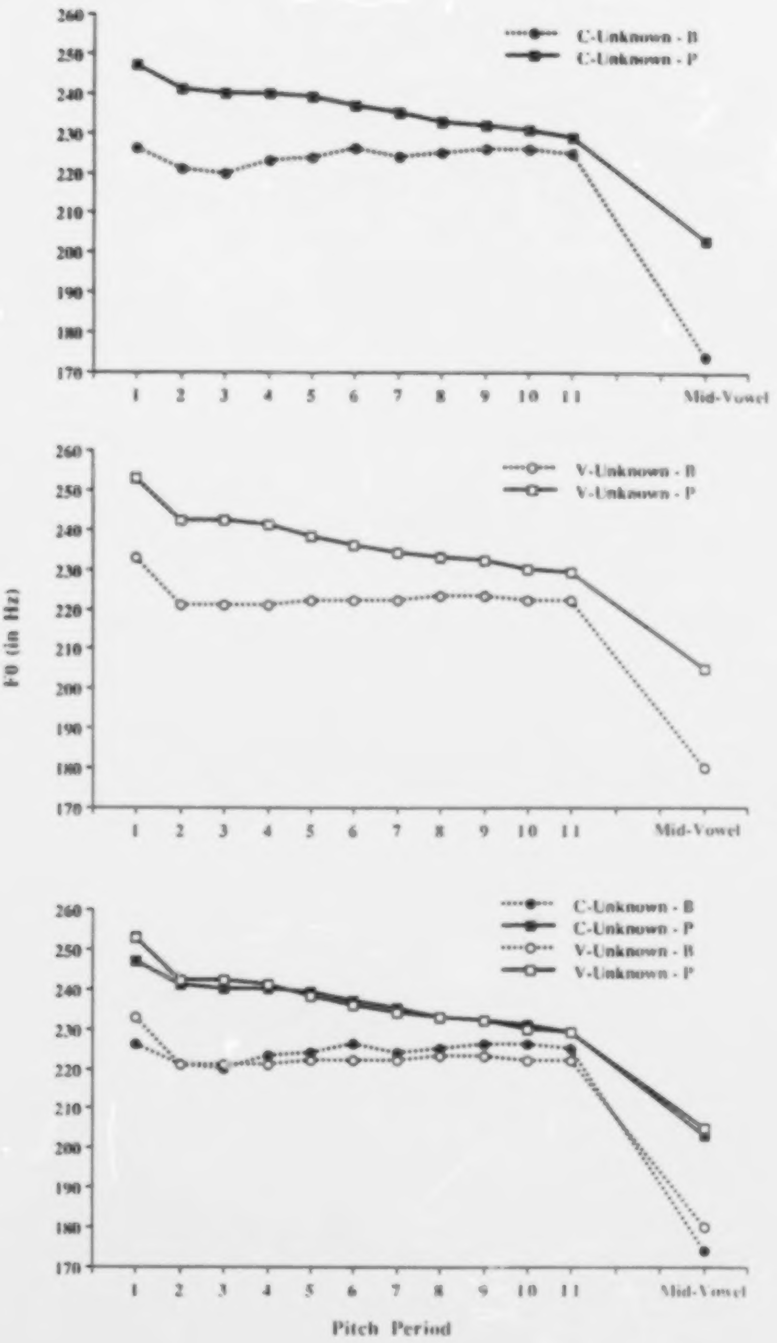


Figure 4. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker DMM, Experiment 1. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

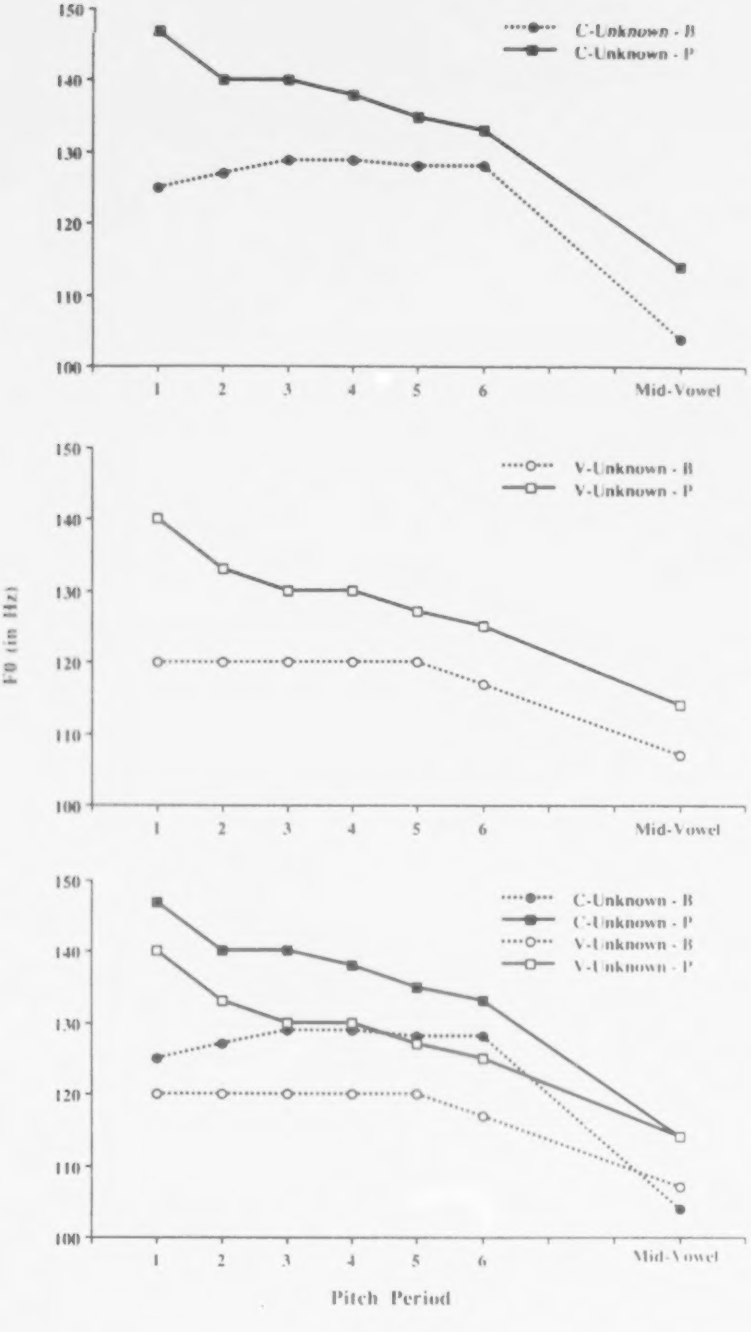


Figure 5. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker LMG, Experiment 1. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

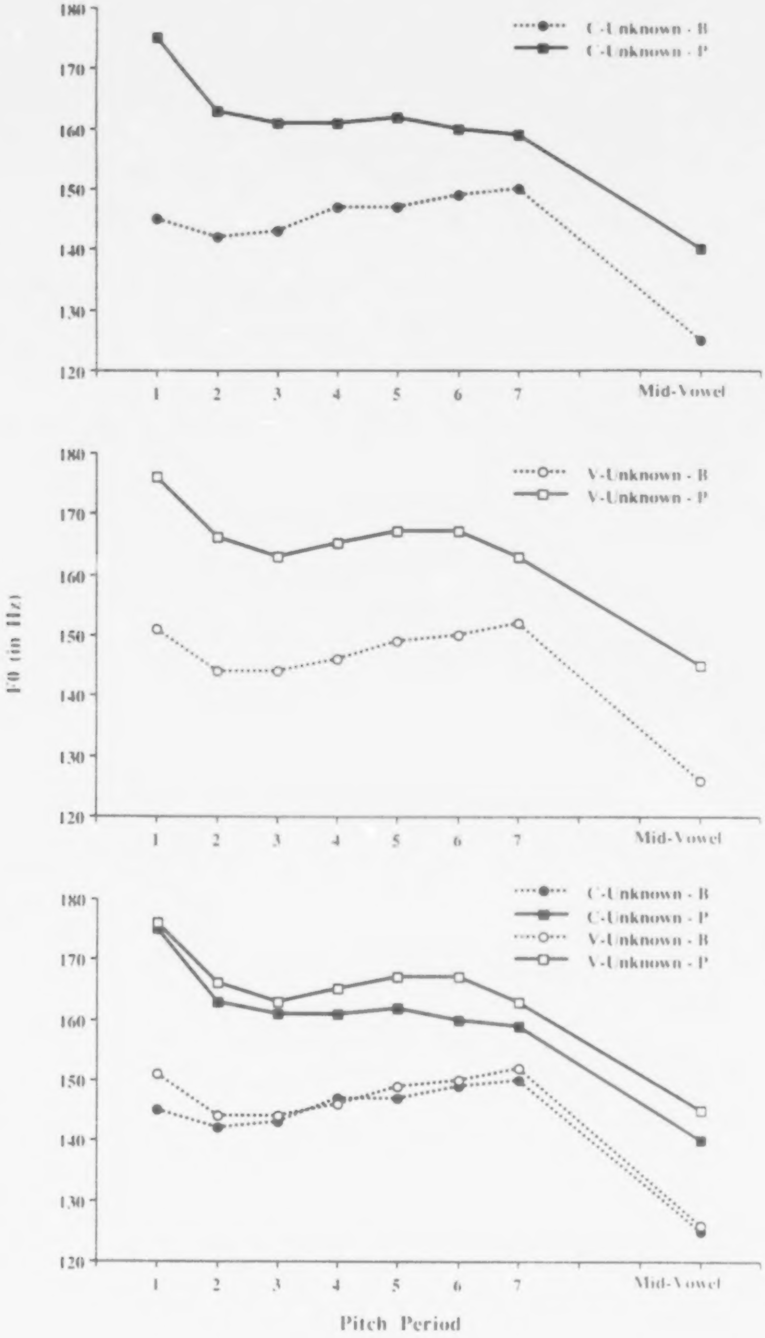


Figure 6. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker DHW, Experiment 1. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

1.3 Discussion

For the three anticipatory effects of coarticulation, there is strong evidence that those effects which are not planned for do not appear, even when there would seem to be ample time for them to be incorporated into the on-going production. This is true both of the frequency domain effects and the time domain one(s). The one perseverative coarticulatory effect did not show a similar tendency. Even though the task for the subjects was unnatural, those segments which were known at the time the utterances were initiated exerted their normal influences. This indicates that the productions were natural to the extent they could be.

Absolute durations for the time between the presentation of the missing letter and the onset of an acoustic effect were in the range defined by a variety of tasks from other researchers. These include a value of 270 ms for fast shadowing (Marslen-Wilson, 1973) and for initiating a learned sequence (Sternberg et al., 1980), and 360 ms for naming one of two letters from print (Cattell, 1886:527; Dick, 1971). It is appropriately longer than the 195 ms (Izdebski & Shipp, 1978) or 260 ms (Watson & Alfonso, 1983) speech initiation time for producing a known, isolated vowel. In another modality, initiation of a series of finger taps has been found to range from 220 to 280 ms, depending on the number of taps (Canic & Franks, 1989). Initiation of incompletely specified finger tapping sequences has been found to be closer to 380 ms (Garcia-Colera & Semjen, 1988; Semjen & Garcia-Colera, 1986, Experiment 3).

The absence of the effect of the upcoming vowel in the F2 of the /a/ in the vowel-unknown condition has an expected component and an unexpected one. At the beginning of the /a/, the speakers would have had to be clairvoyant as well as coarticulatory, since the influencing segment had not even been presented. By the end, however, we might have expected some overlap to occur. The fact that none does occur seems to favor the feature spreading accounts outlined earlier. This conclusion is blunted by the relatively long closure durations obtained. All three subjects had closure durations in excess of the 100 ms window of coarticulation proposed by the frame model (Bell-Berti & Harris, 1981). Two of the subjects had essentially the same closure duration in both conditions (136 versus 138 ms for LMG and 241 versus 247 for DMM for the consonant-unknown versus vowel-unknown conditions respectively). The third subject (DHW) had a longer closure (136 ms) for the vowel-unknown condition than for the consonant-unknown (102 ms). It is possible that the added duration was introduced just because the upcoming vowel had not been properly planned for. The slow rise of the influence of the second vowel may in fact be a necessary aspect of vowel production, according to some overlap models. In that case, the early part of the vowel might be easier to produce during the closure, where there would be no acoustic consequences of the vowel onset (cf. the "hidden" gestures of Browman & Goldstein, in press). Whether this is the case or not, there is no obvious reason within a feature-spreading model for there to be a difference in the closure durations between the two conditions.

To summarize the F2 implications, the lack of an effect at the end of the /a/ is consistent with the prediction of feature-spreading models. But the closure durations are longer than normal in this experiment and particularly so for the vowel-unknown condition, leaving open the possibility that the essential part of the vowel's overlap was hidden by the closure itself. For one subject, the closure was longer in the vowel-unknown condition than in the consonant-unknown, possibly because it is easier to "hide" that part of the vowel than to incorporate coarticulation that had not been planned for.

The effect of the consonant voicing on the duration of the preceding vowel was present in the vowel-unknown condition, but completely absent from the consonant-unknown condition. This seems odd when we consider how much longer the first vowel is in the consonant-unknown condition relative to the vowel-unknown condition. It would not be unreasonable to think that this increased duration would allow the speech apparatus time to introduce the durational differences. Yet, if that was the case, we would still have expected the vowel duration effect to show up in the second half of the vowel, and thus perhaps be of the same magnitude but reduced proportion compared with the vowel-unknown condition. Even if the duration effect decreased in magnitude, we still might have expected to see the nearly universal vowel duration effect, which, though smaller, shows up in most languages (Chen, 1970). The lack of an effect on closure durations is also indicative of a need for this difference to be planned for. Given the differences in acceleration between voiced and voiceless stops (Summers, 1987), we might have expected that the closure duration difference was a necessary consequence of producing the two kinds of stops. Such is not the case. Here again, if the coarticulatory effects are not planned for, they do not show up.

The difference in the behavior of F1 between the two conditions shows a planning effect as well. In the vowel-unknown condition, where the consonant is known, there is a clear effect on the F1 offset due to the voicing of the stop. When the consonant is not known, this effect disappears almost entirely. Only in the last 20 ms before closure do we see an effect, while the other condition shows an effect at least 60 ms back. The conservative behavior of DMM, who had a period of voicelessness in each of her closures regardless of the voicing of the stop, might have contributed to this lack of a voicing effect. Yet the effect was just the same for the other two subjects, who often maintained voicing throughout the closure for voiced stops. Thus the effect on offset frequency is not due directly to the successful cessation of voicing (cf. Fant, Ishizaka, Lindqvist & Sundberg, 1972). Nor is it a by-product of the duration difference between the two stops, since there was no duration difference in the consonant-unknown condition. Rather, there are differences in the shape of the closing trajectory which are present when they are anticipated and all but absent when they must be incorporated mid-stream.

The voicing of the consonant did not affect F1 throughout the vowel as it had in Summers (1987). Summers' stimuli were quite different, being CVC's in sentential context, which made for shorter durations. So the lack of an effect at the beginning of the vowels in the present experiment might be due simply to the relatively long durations of those vowels. Another possibility is that Summers' F1 effect might, in fact, reflect coarticulation between the two consonants, thus affecting the vowel only indirectly. Stimuli like Summers', which had more typical durations than the ones used here, but which were vowel-initial or perhaps /h/ initial would help decide this question.

For the most part, the carry-over articulatory effect (the F0 pattern due to the voicing of the stop) does not show an effect of initial planning. All three subjects have a higher F0 for the voiceless stop than for the voiced at the onset of voicing after release, and at a point well beyond the release of the stop. This indicates either that both the early and late effects of voicing on the vowel's F0 are an automatic consequence of involving the cricothyroid muscle in the devoicing gesture (as argued by Löfqvist, Baer, McGarr, & Story, 1989), or that the planning that such an effect requires can be done in the time it takes to complete the first vowel and the stop closure. Since the differences in the present case are well beyond the approximate upper limit on relaxation time of the cricothyroid (about 120 ms; Löfqvist et al., 1989:1320), some further explanation is called for. Whether that explanation is another automatic one (such as laryngeal height) or one that calls for planning remains to be seen.

While we have evidence that subjects can complete the partially-specified utterance they have launched, we do not see the restitution of anticipatory coarticulation effects in that utterance. The perseverative effect, in contrast, shows very little difference between the two conditions. This outcome is compatible with feature-spreading accounts of coarticulation, though there are considerations that make articulatory-overlap accounts acceptable as well. The feature-spreading accounts predict that even interpolation will be absent, since there is no feature to interpolate. However, such accounts have no explanation for the increased closure duration made by DHW in the vowel-unknown condition. The overlap accounts might attribute this added duration to an act of "hiding" the gradual increase of the influence of the second vowel. This would allow the formant structure of the first vowel to proceed as planned, with the second vowel's onset being accomplished in the silent closure. More disturbing for the overlap accounts is the virtual lack of an F1 effect for the consonant-unknown condition. It would seem that success in making a stop would include the modification of this parameter, due most likely to a difference in the speed of the gesture. This difference is almost absent, however, indicating that even overlap accounts must allow either for planning of coarticulation or for different but linguistically equivalent gestures for the stops.

2.0 EXPERIMENT 2

The second experiment was designed to examine two perseverative effects rather than just one, and to put the location of the missing phoneme somewhat farther along in the utterance. If recovery of planning can occur within the span of a vowel and a stop closure, adding an extra syllable should restore almost all of the coarticulatory effects. To this end, the test stimuli contained an initial syllable consisting only of schwa, followed by /biba/, /bipa/, /buba/, or /bupa/. The missing segment was either the first full vowel (/i/ or /u/) or the second consonant (/b/ or /p/). Stress was to be placed on the first full vowel. The influence of the vowel on the F2 of adjacent vowels could be examined both in the schwa and the /a/.

With the /a/, the effect is no longer an anticipatory one, as in the first experiment, but a perseverative one. F2 vowel effects in the schwa should follow the pattern found for /a/ in the first experiment.

2.1 Method

2.1.1 Subjects

The subjects were the two male subjects of the first experiment and a different female colleague (CLS). The conservative strategy adopted by the earlier female subject seemed likely to further submerge the differences in the planning conditions, so she was replaced. All were native speakers of American English.

2.1.2 Stimuli

Each utterance consisted of a trisyllabic nonsense word. The first syllable was the mid-central unstressed vowel /ə/. The second syllable began with /b/, with either /i/ or /u/ as the vowel. The third syllable began with either /b/ or /p/ and ended with the low-central vowel /a/. These were represented on the computer screen as UHBIBA, UHBIPA, UHBUBA, and UHBUPA. Stress was to be placed on the middle syllable.

2.1.3 Equipment and Procedure

Subjects produced the utterances in response to a visual display on a computer screen. In the first condition, the position for the /i/ or /u/ vowel appeared initially as a blank (e.g., "UHB__PA"). When this display appeared, the subject could take as long as he or she wished to prepare before initiating a response. Once the subject began phonating the /ə/, the missing letter (I or U) immediately appeared in the previously blank position. The subject was instructed to incorporate that vowel into the ongoing production as rapidly and as smoothly as possible. In the second condition, the position for the consonant appeared initially as a blank (e.g., "UHBU__A"), and the rest proceeded as in the other condition.

Twenty repetitions of each utterance type were presented in random order. Errors were again excluded from the analysis. The utterances were input into the computer as in the previous experiment and measured as before, with duration and F0 measured by hand from the waveform and F1 and F2 taken from the LPC analysis.

2.2 Results

The same coarticulatory effects which were examined in the first experiment were looked at here. The main differences were that the F2 effect of the /i/ and /u/ vowels could be seen both in anticipatory position on the schwa and in perseverative position on the /a/. The duration effect associated with the voicing of the medial /b/ or /p/ now appeared on the /i/ and /u/ vowels rather than the /a/, as did the F1 effect.

Error rates (which included mispronunciations of either the vowel or the consonant) for the three subjects for the consonant-unknown and vowel-unknown conditions were: 11.3% and 17.5% (CLS); 7.5% and 20.0% (LMG); and 7.5% for both for DHW. All errors in the consonant-unknown condition were consonant mistakes, while vowel mistakes accounted for only 42% of the errors in the vowel-unknown condition. This pattern differs somewhat from that in the first experiment, where neither condition had an effect on the type of error.

Acoustic consequences of the missing phonemes for the consonant-unknown condition occurred at about the same point as they did in the first experiment, at around 320 ms. For the vowel-unknown condition, evidence appeared earlier, at around 230 ms.

Figure 7 shows the effect of the middle vowel, /i/ or /u/, on the F2 of the initial schwa, measured at the beginning, the midpoint, and just before closure. F2 of schwa is affected throughout its duration when the vowel is known at initiation time, but is affected only the end when the vowel is not known. The effect of vowel identity on F2 was strong ($F(1,398) = 10.37, 48.01, 60.05, p < .01$ for the beginning, middle and end respectively), though it did differ by condition ($F(1,398) = 12.48, 20.97, 13.60, p < .001$ for the interaction of condition and vowel for beginning, middle and end respectively). In the vowel-unknown condition, the beginning ($F(1,192) < 1.0$) and mid point ($F(1,192) = 2.65$, n.s.) were unaffected by the upcoming vowel, while the end was ($F(1,192) = 4.93, p < .05$). Thus, in contrast to the first experiment, the final portion of the vowel was affected even in the vowel-unknown condition, though to a far smaller degree than in the consonant-unknown condition. Note also that the amount of vowel-to-vowel coarticulation increases (from 29 Hz to 107 Hz) as closure is approached in the consonant-unknown condition, but varies unsystematically within a small range (7-28 Hz) for the vowel-unknown condition.

It is of interest that the closure for the /b/ between the schwa and the middle vowel was not abnormally long for one of the subjects (DHW) in this condition, although his closures had been unusually long in the first experiment in this condition. In fact, for DHW, the closures had essentially the same duration in both conditions.

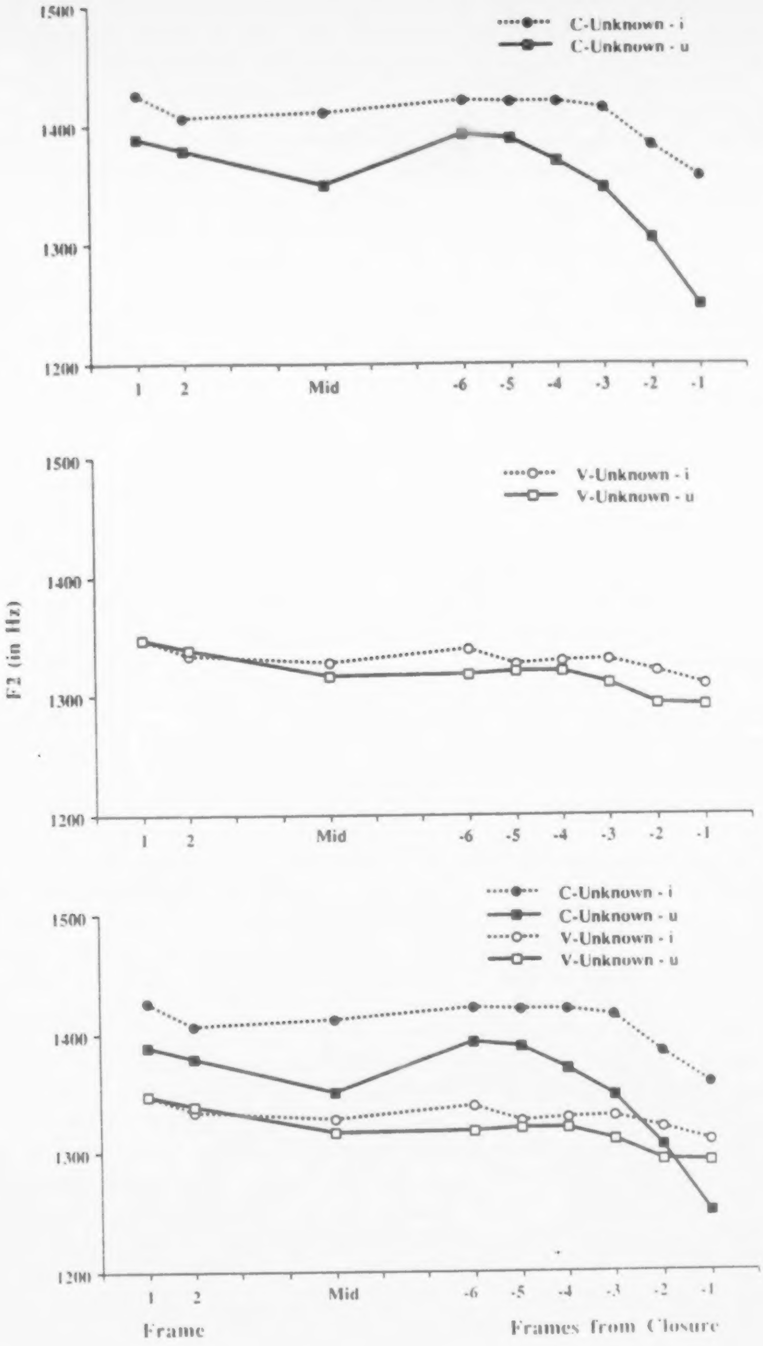


Figure 7. **Effect of an upcoming vowel on the F2 of initial /ə/, Experiment 2. In the consonant-unknown condition (top panel), the vowel is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is not. The lower panel superimposes the upper two panels. The first two points are the analysis frames at the beginning of the vowel. The next point is the frame at the mid-point of the /ə/ as defined by its duration. The last points are the frames just before closure.**

For the other two subjects, this /b/ closure was longer in the vowel-unknown condition than in the consonant-unknown. Given that one explanation of the lack of an effect of the upcoming vowel on the F2 in Experiment 1 was that the closure was effectively "hiding" the vowel gesture, a more direct test of this possibility here seemed in order. The correlation between closure duration and the size of the effect on F2 of the following vowel was measured for each subject for each condition. To obtain the size of the effect, the mean for the other vowel was subtracted from each F2 value. More specifically, for the vowels in the /i/ context, the mean value for each frame of the schwa transition in /əbuba/ was subtracted from the corresponding frame of the means of the /əbiba/ transitions. Similarly, /əbupa/ was subtracted from the mean of the /əbipa/ tokens. For /u/, the effect was in the other direction, so the frames of the /əbuba/ transitions were themselves subtracted from the mean value of the /əbiba/ frames, and the /əbupa/ from the /əbipa/. In this way, effects in the predicted direction were always positive, while negative numbers reflected the opposite of the predictions. Average values for the last three frames before closure were then correlated with the closure duration of the first stop, /b/. (For both the F2 differences and the durations, values outside two standard deviations from the mean were excluded.) A negative correlation would show that the shorter the closure is, the greater the effect of the up-coming vowel.

As shown in the correlation values of Table 1, two of the three subjects showed a significant negative correlation in the last three analysis frames of the schwa in the vowel-unknown condition. No significant correlations occurred in the consonant-unknown condition, but this may be due to the small standard deviations of the closures (see column 5 of Table 1), leaving little room for a correlation to appear. The subject who had a correlation that just missed being significant (DHW) had the only significant vowel-to-vowel effect in individual analyses. (These individual analyses were performed in conjunction with the correlation analysis, and were done in spite of the fact that there was no interaction between subject and vowel effect in the main analysis.) Here again, the standard deviation of the closure durations was small. Thus it appears that when the vowel has not been planned for, it can affect the end of the preceding vowel as long as it is not "hidden" by the stop closure.

Table 1. *Correlations between the size of the effect of upcoming F2 (averaged over the last three analysis frames of the schwa vowel) with the closure duration for the stop.*

Subj	Condition	Pearson's r	N	$p <$	Closure	SD
CLS	V-Unknown	-.285	61	.05	140.9	47.2
	C-Unknown	.074	65	n.s.	69.3	11.5
LMG	V-Unknown	-.336	59	.01	89.5	25.5
	C-Unknown	-.023	72	n.s.	82.9	11.1
DHW	V-Unknown	-.190	67	(.15)	61.8	14.3
	C-Unknown	.103	70	n.s.	64.6	8.1

As can be seen in Figure 8, the effect of the preceding vowel on the F2 of /a/ exists well into the vowel, and equally in the two conditions. An analysis of variance for the first fifteen analysis frames with the grouping factors of speaker, condition, consonant, and vowel was performed. A large number of frames was used so that at least half of the comparisons for /p/ would involve voiced formants. (The mean VOT for /p/ was 61 ms, so the first six frames would contain no voiced formants, and aspiration shifts the frequency of the formants; see Repp & Lin, 1987.) The effect of the preceding vowel was quite strong ($F(1,397) = 98.07, p < .001$). There was no interaction of vowel and condition, with or without other factors. Thus the effect of preceding vowel on F2 was present whether that vowel had been known at initiation time or not, and to an equivalent degree in both conditions.

One more effect of F2 which has been postulated to be coarticulatory by some (e.g., Fant, 1973), namely, the higher F2 for aspirated stops compared with unaspirated, can be examined here as well. For both conditions, F2 was about 200 Hz higher for /p/ than for /b/ at the consonant release (Figure 9), an effect which matches the results of Repp and Lin (1987) almost exactly. This supports their conclusion that the effect is an acoustic one, dependent only on the effects of coupling due to the open glottis. Unlike their results, however, there was not a complete convergence of the F2 once the vowel steady-state was reached. Instead, the steady-state F2 after /p/ was actually about 35 Hz lower than that after /b/. Perhaps this is an over-compensation of some sort, but it does not vary with condition, so it will not be further explored.

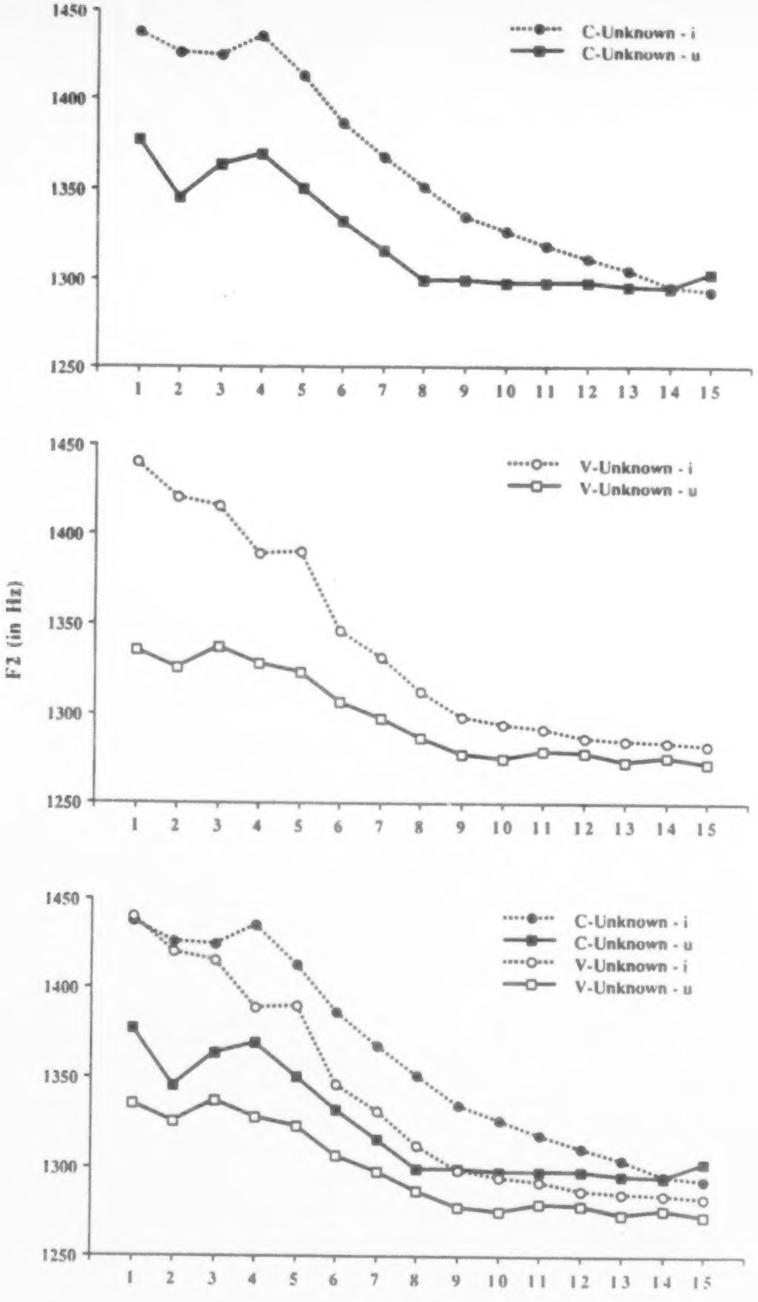


Figure 8. **Effect of an preceding vowel on the F2 of final /a/, Experiment 2. In the consonant-unknown condition (top panel), the vowel is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is not. The lower panel superimposes the upper two panels. Each frame represents 10 ms of speech.**

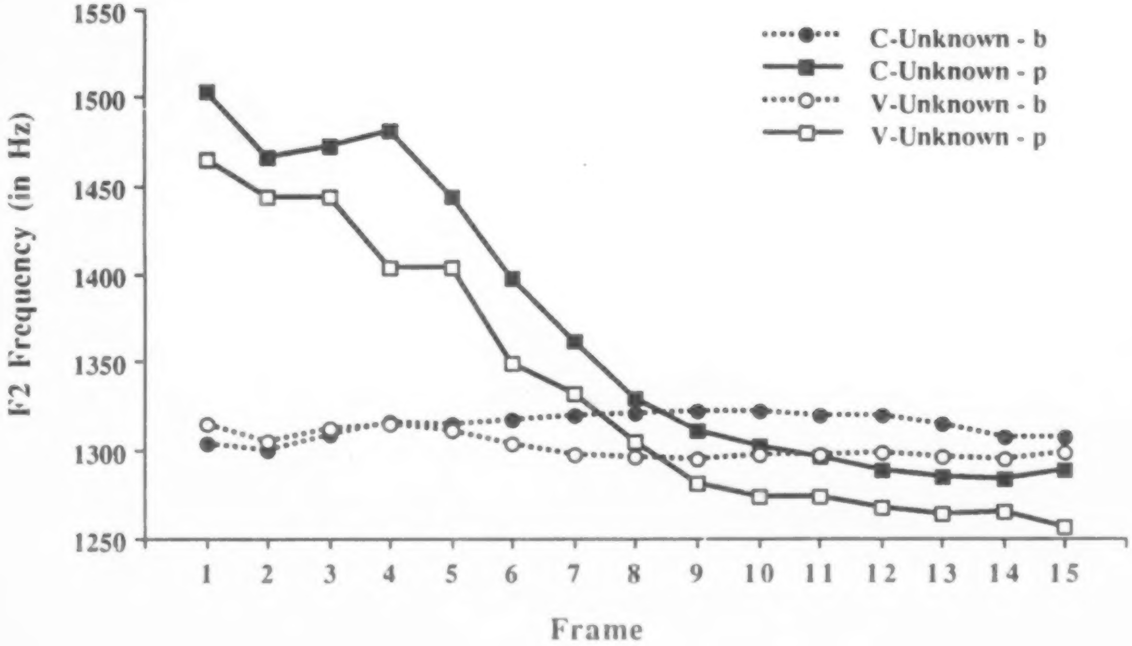


***Figure 9*. Effect of the voicing of a preceding consonant on the F2 of final /a/, Experiment 2. In the consonant-unknown condition (top panel), the consonant is not known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each frame represents 10 ms of speech.**

The effect of the consonant voicing on vowel and closure duration can be seen in Figure 10. The bars in this figure have been aligned at the onset of the middle vowel, thus showing the differences in the duration of the first syllable yet allowing the voicing effect to be seen more easily. In the consonant-unknown condition, voicing affects neither the vowel duration (164 ms before /b/, 159 ms before /p/, $F(1,205) < 1.0$, n.s.) nor the closure duration (103 ms for /b/, 102 ms for /p/, $F(1,205) < 1.0$, n.s.). When the consonant is known (vowel-unknown condition), by contrast, both the vowel duration (120 ms before /b/, 114 ms before /p/, $F(1,192) = 8.41$, $p < .01$) and the closure duration (74 ms for /b/, 91 ms for /p/, $F(1,192) = 35.79$, $p < .001$) are influenced.

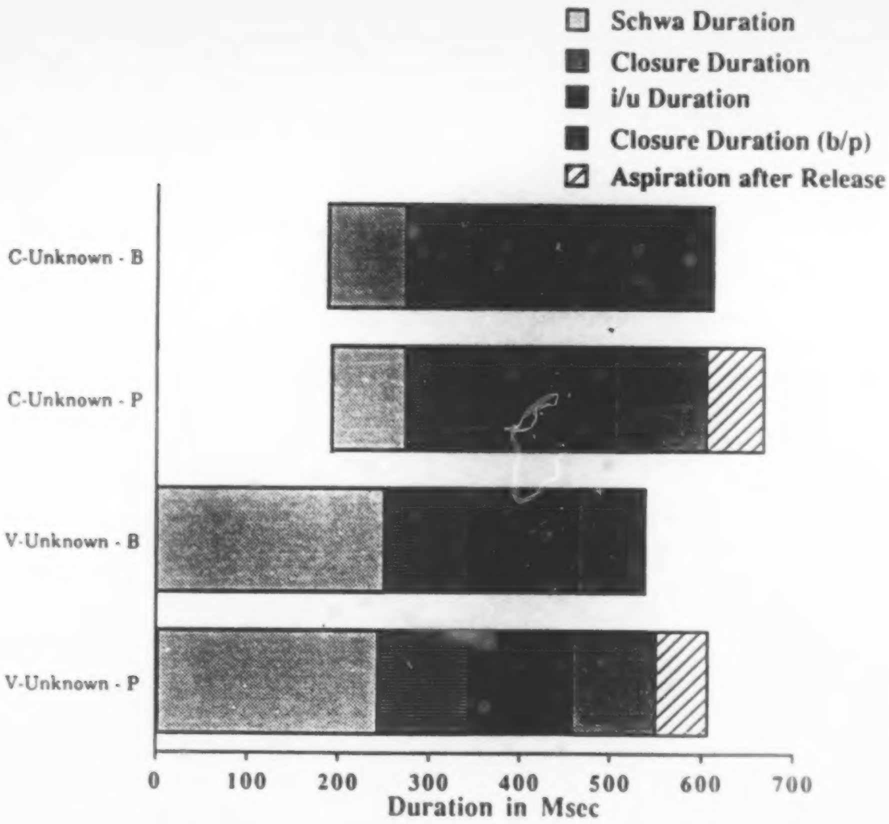


Figure 10. **Effect the voicing of an upcoming consonant on the duration of initial /a/ and the stop closure, Experiment 2. In the consonant-unknown condition (top two bars), the consonant is not known at utterance initiation time. In the vowel-unknown condition (lower two bars), it is.**

The F1 values for the offset of the /i/ or /u/ vowel are displayed in Figure 11. They show an extremely small trend toward a lowering due to voicing of the stop. Statistical analysis shows that this is not significant either as a main effect of voicing ($F(1,398) < 1.0$, n.s.) or as an interaction with condition ($F(1,398) < 1.0$, n.s.). For the final frame by itself, there is a marginal effect ($F(1,398) = 3.49$, $p < .10$). In short, F1 was unaffected by the voicing of the stop whether it was the consonant which was unknown or the vowel.

Figures 12 to 14 show the effect of voicing on the F0 after release of the consonant, again for each subject separately, to overcome the difference in the number of data points. For speaker CLS (Figure 12), there are main effects of condition and voicing ($F(1,128) = 5.85$ and 6.21, $p < .05$ respectively) with no interaction ($F(1,128) = 1.78$, n.s.) For speaker LMG (Figure 13), there is an effect of voicing for both the initial portion in the consonant-unknown condition ($F(1,69) = 24.36$, $p < .001$), and in the middle ($F(1,69) = 15.26$, $p < .001$). There was no effect at the beginning in the vowel-unknown condition ($F(1,57) = 1.04$, n.s.) though there was in the middle ($F(1,57) = 7.51$, $p < .01$). Though there is a difference between conditions for this subject, it is not a difference that could be predicted by the nature of the missing phoneme, since the consonant effect on F0 failed to appear in the condition where the vowel was unknown. For speaker DHW (Figure 14), the F0 results were similar to those of Experiment 1. The pitch periods immediately after release differed in the expected direction for both the consonant-unknown condition ($F(1,70) = 44.93$, $p < .001$) and the vowel-unknown condition ($F(1,70) = 137.64$, $p < .001$). The same held true of the mid-vowel pitch periods ($F(1,70) = 4.91$, $p < .05$, for the consonant-unknown condition, and $F(1,70) = 57.09$, $p < .001$, for the vowel-unknown condition).

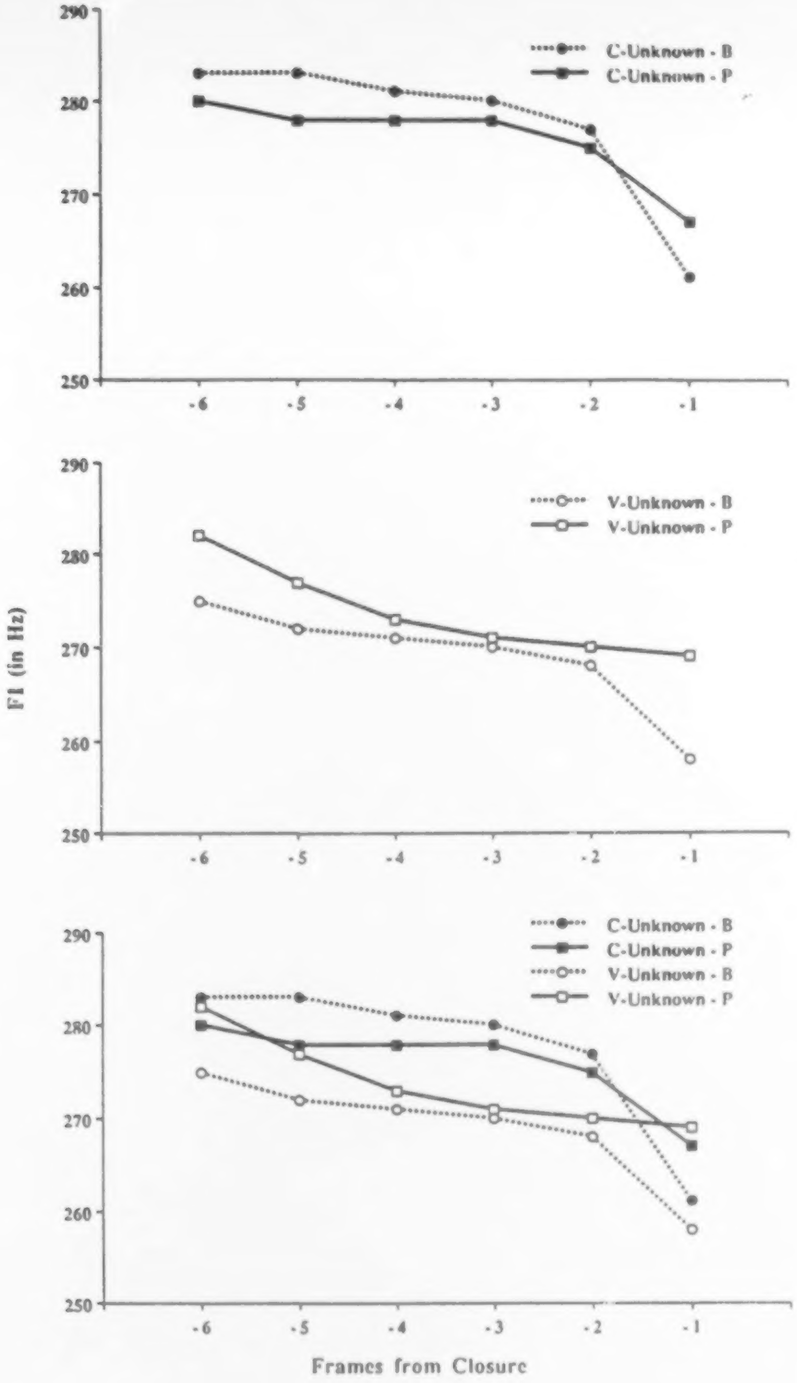


Figure 11. **Effect the voicing of an upcoming consonant on the F1 of medial /i/ or /u/,** Experiment 2. In the consonant-**unknown condition (top panel), the consonant is known at utterance initiation time.** In the vowel-unknown condition **(middle panel), it is. The lower panel superimposes the upper two panels.**

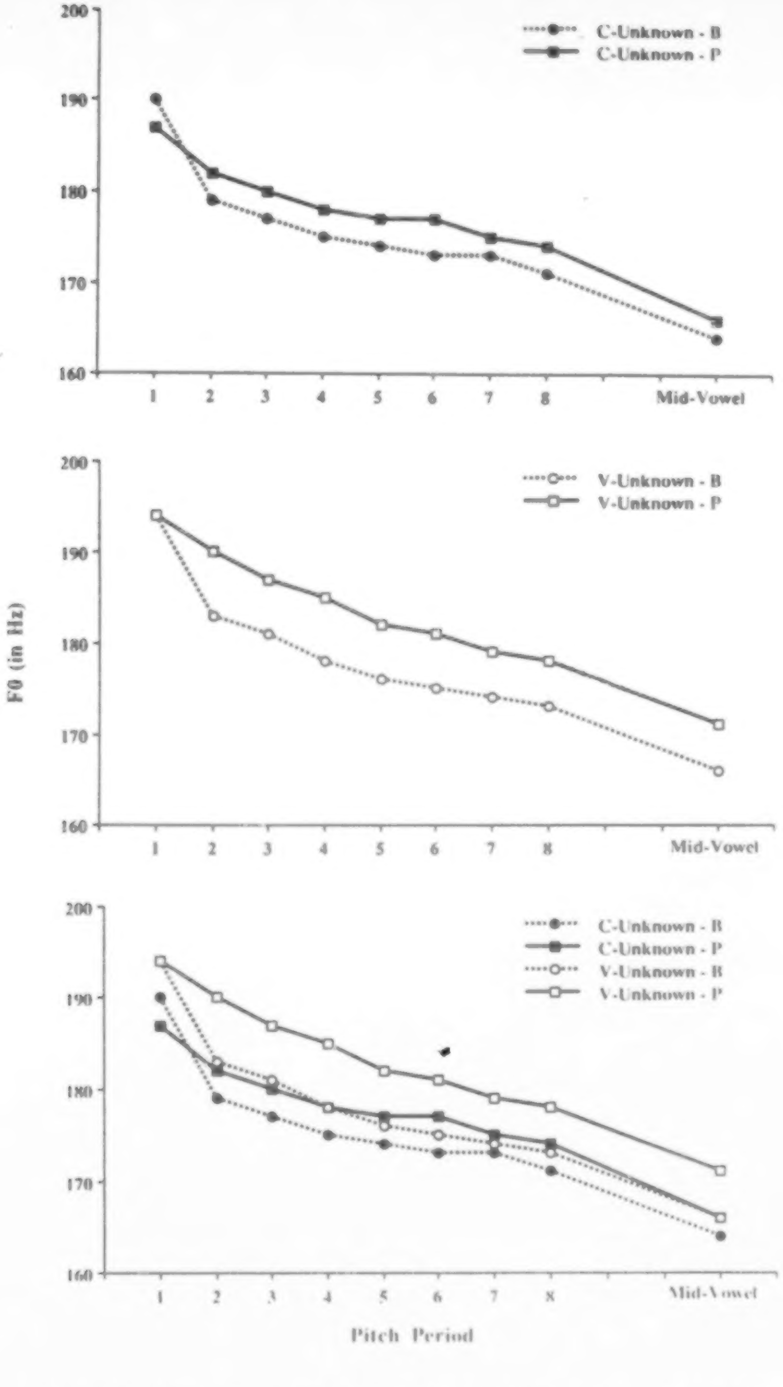


Figure 12. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker CLS, Experiment 2. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

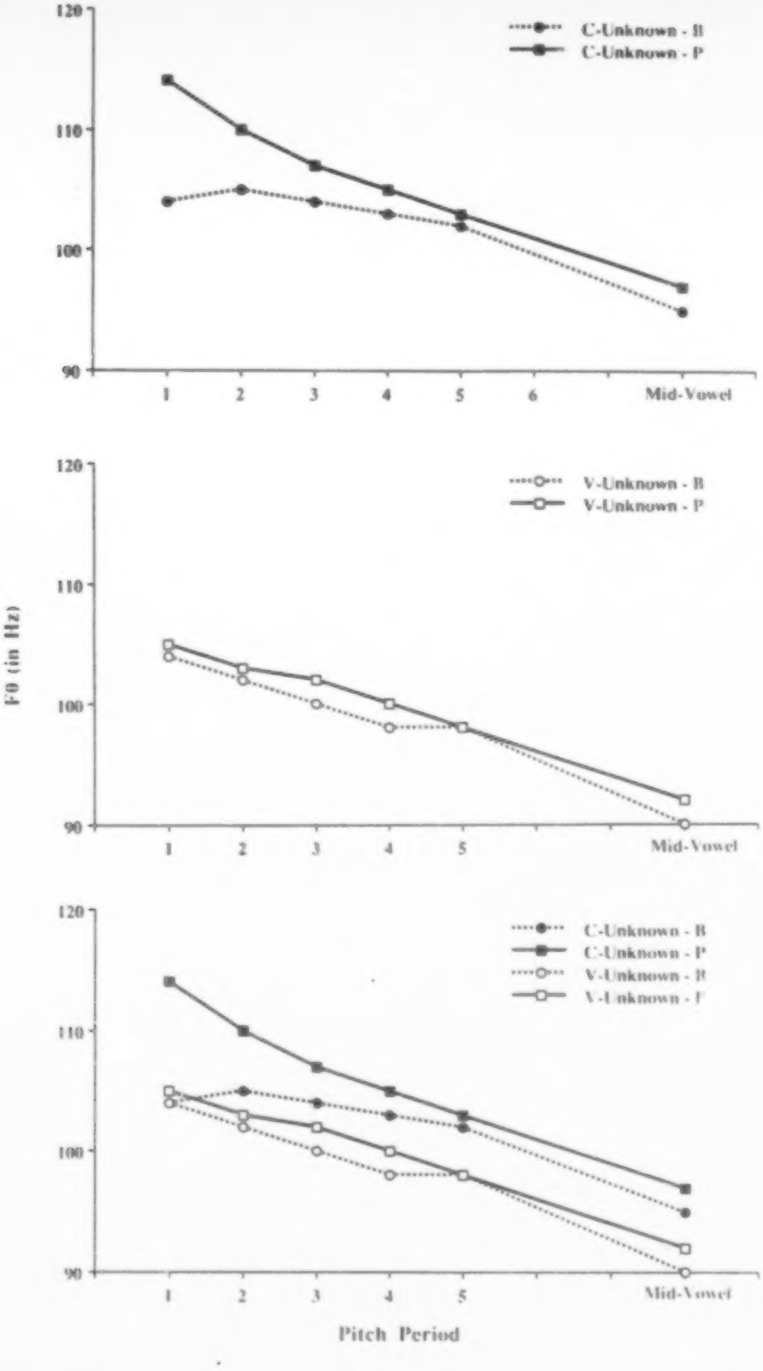


Figure 13. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker LMG, Experiment 2. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

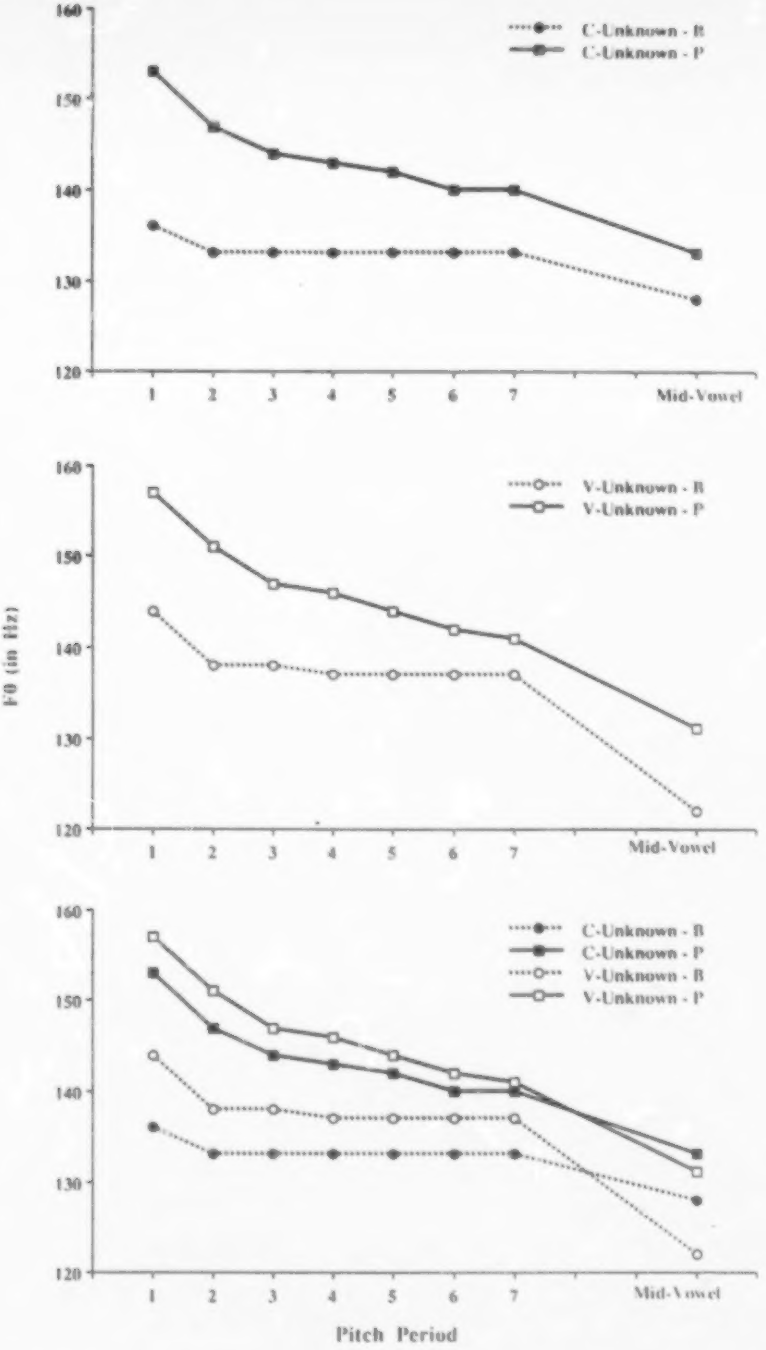


Figure 14. **Effect of the voicing a preceding stop on the F0 of final /i/ or /u/ for speaker DHW, Experiment 2. In the consonant-unknown condition (top panel), the consonant is known at utterance initiation time. In the vowel-unknown condition (middle panel), it is. The lower panel superimposes the upper two panels. Each pitch period was measured individually, so the higher F0's are of shorter duration.**

2.3 Discussion

Of the three anticipatory effects of coarticulation measured here, two (the F2 effect of the up-coming vowel on schwa and the duration difference due to voicing of the stop) showed the predicted difference between the conditions, while the third (the effect of voicing on F1) failed to show up in either condition. This latter failure seems to be due to the intrinsically low F1 of the vowels involved. All three of the perseverative effects (the effect of the preceding vowel on the F2 of /a/, and the effects of voicing on the F2 and F0 of /a/) showed no difference between the conditions. The conclusions from the first experiment receive further support, that coarticulation is largely planned, and that some of this planning can occur during the course of an utterance. The effect of the i/u vowel on the F2 of the preceding schwa does show some partial restoration under favorable conditions in the vowel-unknown condition, providing support for an overlap model.

The vowel-to-vowel coarticulation showed up on initial schwa in much the same way that it appeared on the /a/ in the first experiment, with one interesting difference: Just before closure, the vowel effect began to appear even in the vowel-unknown condition. Further, there was a correlation for two of the subjects between the duration of the closure separating the schwa from the /i/ or /u/ and the size of the coarticulatory effect. When the closure was small, the effect was large, showing that articulatory overlap was having acoustic consequences even though the vowel had not been planned for before utterance onset. Such evidence is consistent with the notion that the articulatory onset of vowels precedes the typical acoustic landmarks by a considerable amount (Fowler, 1979, 1980). The largest stretch of this vowel, however, shows no coarticulation.

The perseverative effects of the middle vowel on the final /a/ were present in both conditions, and equally strongly in both. This shows either that perseverative coarticulation is basically mechanical (cf. Recasens, 1984, 1987), or that planning can be restored in the course of the utterance. Differences in magnitude of the effect do not help us decide the question in this instance, since two different vowels were involved. The anticipatory effect was much larger, but schwa is more amenable to this kind of coarticulation than /a/ is. There is no way within this paradigm to sort out these two possible explanations for the restoration of perseverative effects. However, the presence of F0 coarticulation for stop voicing late in the vowel argues that at least some perseverative effects are due to planning rather than to mechanical considerations.

The effects of stop voicing on vowel and closure duration mirrored those of the first experiment. If the consonant was not known at initiation time, the duration effects did not appear. This is in spite of the fact that the syllable which would have exhibited the duration effects was no longer the first one. In this instance, the time and/or the structure of the first syllable was not appropriate for the recovery of coarticulatory planning. (If the unstressed syllable was structurally part of the following stressed syllable, the lack of recovery may have been due to the unavailability of the structure rather than to a lack of time.)

The lack of the effect of voicing on F1 is probably artifactual. High vowels (in this case, /i/ and /u/) have relatively low frequency F1's. Other researchers have found that this obscured the coarticulatory effect (Hillenbrand, Ingrisano, Smith, & Flege, 1984). Thus this portion of the experiment seems to reflect a lack of an appropriate measure rather than a lack of coarticulation.

The behavior of F0 after the release of the stops indicates that the subjects were introducing a typical devoicing gesture (Löfqvist et al., 1989). Though smaller in magnitude than in the first experiment, the effect was consistently present. (The difference in magnitude probably reflects the shifting of primary stress from the syllable containing the stop (Experiment 1) to the preceding syllable (Experiment 2).) The effect in the midpoint of the vowel, however, does not seem likely to be due to the cricothyroid activity for the voicelessness per se, since the duration of the effect is so long. Some other automatic effect (such as laryngeal height) may be responsible. If not, then we may have evidence that planning has been recovered over the course of the utterance, rather than an indication of a difference between anticipatory and perseverative coarticulation.

3.0 GENERAL DISCUSSION AND CONCLUSION

Coarticulation, though presumed to be due to the constraints of producing speech in real time, is largely a result of planning an utterance rather than an automatic consequence of successfully producing that utterance. This was shown by forcing subjects to begin uttering a nonsense string before the entirety of the string was known. The missing element was either a vowel or a consonant, and, for the most part, anticipatory coarticulatory effects associated with the missing segment did not appear even though effects associated with the known segment did. Subjects were, therefore, successfully producing a normal utterance to the extent possible, and inserting a new phoneme as the experiment required. Perseverative effects, however, were mostly present in both conditions. This indicates either that perseverative effects are indeed different from anticipatory ones in being mechanical rather than planned, or that planning for this effect can be accomplished after utterance onset. The relatively long duration of the F0 effect due to voicing argues for a planning explanation. In the second experiment, the appearance of vowel-to-vowel coarticulation in just those cases where a stop closure was short enough for articulatory overlap to have an acoustic effect indicates that some coarticulation is indeed automatic. If coarticulation consisted only of such effects, however, its description would be relatively simple. The effects that cover a much larger stretch of time, the ones which need to be planned, are the ones that make coarticulation such a challenging problem.

Under the best of circumstances, these results would allow us to test the differing predictions of feature-spreading and articulatory overlap models. Unfortunately, there is little agreement in phonology on which features are specified for particular values and thus which features might spread. In addition, feature-spreading accounts have been modified in recent publications, while overlap models in general have not. For example, Keating (1985) proposes interpolation as an adjunct to feature spreading. Thus the best we can do is to elaborate some of the implications for each class of theories of the results for the present study. The descriptions of feature-spreading accounts, though, will usually be of modified versions. As we will see, there will remain difficulties for the feature-spreading accounts, and necessary amendments for the overlap accounts.

If vowel-to-vowel coarticulation is based on phonological features, then the degree of such coarticulation should be predictable from the features used by a language. With a complex vowel system using many features, we could expect that there would be less coarticulation than with a simple vowel system which leaves many features unspecified. In fact, that is what we find. When a language has a small vowel inventory, it seems to allow greater coarticulation (Magen, 1984; Manuel, 1987; Manuel & Krakow, 1984). These authors propose explanations which do not depend on features, however. Manuel (1987), for example, argues that speakers are sensitive to the acoustic differentiation of the vowels of their language, so that small inventories allow greater latitude for subphonemic variations. That is, she predicts that vowel-to-vowel coarticulation will be more globally related to a vowel's articulation rather than to the phonological features, but with output constraints that force it into a particular part of the vowel space.

Perhaps the most supportive aspect of the current study for the feature-spreading accounts is that some of the coarticulation was small and evenly distributed throughout the vowel affected. For example, the effect of the up-coming vowel on the F2 of /a/ in Experiment 1 (Figure 1) is of the same magnitude at the beginning as it is up until the last frame before closure. If this coarticulation is due to the spread of a phonological feature, such consistency is to be expected, although the "mechanical" effects at the end of the vowel can also be expected. Certainly some dynamic overlap models will have to be modified to account for such effects, but the types of interpolation used in feature-spreading models must also be constrained if they are to allow for relatively steady-state effects. However, the type and size of coarticulation must also be predicted by the features used if the feature-spreading models are to be fully supported. The effects reported here seem too small to be due to the implementation of a feature such as [+round]. In the example here, there is some evidence that the effect in the consonant-unknown condition is due to the /i/ context, since the /u/ context and the lack of an effect are similar (see the bottom panel of Figure 1). If so, then the choice of the feature to account for the coarticulation is difficult. It is surely not height, which would show up on F1, not on F2, and which is already contrarily specified for /a/. The effects on F2 will be due to front/back and/or rounding. Yet /a/ cannot change its specification for "back" without becoming a different vowel, and it is already "-round" (possibly distinctively). The remaining features available in a feature-based vowel system, labial and ATR, do not seem to do the trick. Even if they do, the size of the effect obtained (in this case, around 25 Hz) seems much too small to be attributed to the change of a feature. It may indeed be important, as Keating (1988) proposes, to allow feature matrices to exit the phonology with some features unspecified, but relying on the inventory of phonological features does not seem to give enough detail to explain coarticulation. The same can be said of any feature-spreading account which does not have more (or at least more continuous) phonetic features than phonological ones.

Theories of coarticulatory overlap have different difficulties to handle. The most obvious, both from the present work and that of Magen (1989), is that overlap must extend beyond the stressed vowel (indeed, beyond the stress foot in Magen's case). This overlap must be extensive in time but not in magnitude. Since the size of the effect does not vary through the majority of the vowel, simply interpolation algorithms will not serve here any better than they do with feature-spreading accounts. An extension of Articulatory Phonology (Browman & Goldstein, 1985, 1986) might allow for the modification of one vowel's input parameters as well as articulatory overlap with the same influencing vowel. This two-component approach (i.e., both gestural modification and gestural overlap are allowed) fits the present results (especially those of Experiment 2) quite well, but it remains to be seen whether the increase in the power of the model is warranted. An alternative can be put in terms of prominence curves (Fowler & Smith, 1986), which do not correspond directly to gestures (since prominence is associated with phonemes, not gestures) or to physiological measurements but are rather theoretical constructs which are intended to express, nonmathematically, the structure of utterances. The range of phenomena addressed by prominence curves would be enlarged by the postulation of extended "tails", which would allow the overlap of vowels at a reduced level for long periods. While it may be possible to make Articulatory Phonology and the prominence curve account generate the same results, two aspects of coarticulation suggest that the "tail" approach might be more natural. First, having the tail be a part of the prominence curve makes it natural to assume that the characteristics of the tail depend on the characteristics of the main prominence. With parameter modification, such a relationship can be specified but need not be. Second, if the tail is indeed connected to the main prominence, then the connection should be maintained for the tail to exist. If, for example, a consonantal gesture that was incompatible with the vowel's tail preceded the vowel, it would be reasonable to think that the tail would be suppressed. This would begin to account for the apparent lack of coarticulation across consonants in Russian (Keating, 1988).

One result from each experiment supports the gestural overlap approach. In Experiment 1, one subject (DHW) had longer closure durations in the Vowel Unknown condition, perhaps in response to the greater ease of hiding the onset of the vowel gesture in the silent stop closure than superimposing it on the end of the already launched preceding vowel gesture. In Experiment 2, the closure durations of the first stop were short enough that correlations could be made between the size of the vowel-to-vowel coarticulation and the closure duration. In the Vowel Unknown conditions, there was indeed a negative correlation, indicating that when the closure was short, the coarticulatory effect was big, just as the overlap models predict. Since the overlap models were originally designed to show that the vowel gesture necessarily precedes the point usually measured as "vowel onset" (Fowler, 1979), these results are promising. Even though modifications need to be made, it indeed seems that the a failure to produce the "tails" (or equivalent) would still allow the phoneme to be recognized correctly, while the effects near the stop closure are unavoidable because they represent the true onset of vowel activity.

The results for the perseverative coarticulatory effects are somewhat ambiguous. Since they appear, for the most part, equally strongly in both conditions, it is either the case that they are successfully planned for during the preceding vowel and stop closure, or that they do in fact represent physical constraints as claimed by such researchers as Gay (1977) and Recasens (1987). The F0 difference late in the vowel is the one most difficult to attribute to mechanical effects. The mechanism most often cited to explain the F0 difference, the residual effect of the devoicing activity on the laryngeal control of F0, does not seem to extend long enough to account for the effects found here and elsewhere (Ohde, 1984; Löfqvist et al. 1989). While other more automatic components may yet be found, it currently seems more likely that even in the perseverative direction, planning is at work. Even the effects attributable to more sluggish articulators may call for a planning explanation. The stimuli used by Magen (1989) and Recasens (1989) (which included a stressless schwa between two stressed vowels), for example, have more distance (in time and segments) between the influencing segment and the influenced than could be encompassed by mechanical effects. Still, they find vowel-to-vowel effects. To that extent, we can lean toward the planning explanation. That is, while there are acknowledged differences in the magnitude and behavior of anticipatory and perseverative effects, we have not exhausted the range of alternative explanations. One promising explanation is that a given amount of coarticulation might affect preceding and following contexts differently because the contexts are themselves inherently different. Recent work by Fowler and Vatikiotis-Bateson (in preparation) showing that amplitude of the speech signal and indeed the speech gestures regularly declines over the course of an utterance (much as F0 does) sets up an interesting possibility: If the strength of a vowel's coarticulation is dependent on the vowel's own strength, then vowels will have a larger effect on following segments than on preceding ones. That is, the strength of the coarticulation would be the same, but the strength of the affected phoneme would be less, resulting in greater evidence of coarticulation. There are, of course, other differences to be explained, but this example shows that the explanations need not depend on different mechanisms for anticipatory and perseverative coarticulation.

The lack of the duration effects of consonant voicing in the consonant-unknown condition may not be surprising given that the effect in English is thought to be phonologized. We should not expect phonological effects to appear by virtue of articulation per se. However, the difference in speed of closure, as seen in the F1 transition differences, is almost absent as well. Both the normal duration and F1 effects seem to be "enhancing" the voicing distinction, as Stevens and Keyser (1989) might say. Yet here again, the two effects do not seem to be easily attributed to specific, binary values of phonological features, as these authors require. As an alternative, the structures and relationships proposed by the task-dynamic approach to speech production (e.g. Saltzman, 1986; Saltzman & Kelso, 1987) may account for the success of producing the voicing distinction even in the absence of many of its normal characteristics. In that approach, the definition of a linguistically significant gesture is not the variety of articulatory movements which go into the utterance, but rather the goal which large family of movements which can attain. Thus if some portion of an articulator's trajectory is disturbed, other articulatory maneuvers can make up the difference. Defining speech as a set of coordinative structures, then, seems to allow the continuous variation and responsiveness to a changing environment needed for describing such facts as those presented here (cf. Saltzman & Munhall, in press).

The magnitude and extent of coarticulatory effects seem to exceed the requirements of the speech articulators for accommodation of the competing demands made by successive phonemes. Acceptable, distinctive utterances can be accomplished without many of the coarticulatory effects found consistently in speech. It seems, then, that coarticulation is geared toward making speaking easier and/or enhancing the effectiveness of the speech stream for the listener. When phonemes influence one another, the cohesion of the sound is presumably increased. We have evidence in reaction times that this information is taken into account even when it does not affect overt phonemic decisions (Martin & Bunnell, 1982; Whalen, 1984). Since these reaction time studies found effects even with single tokens of utterances, it should be the case that even the vowel-to-vowel effects examined here should be usable, despite the large number of repetitions required to measure the effect. Fowler and Smith (1986), in fact, found that vowel-to-vowel coarticulation does affect perception, though they examined the effects on schwa. Acoustic measurements seem less sensitive than a human listener, so a listener's ability to detect the presence of this kind of coarticulation should not be ruled out simply because we cannot measure it successfully on single tokens. The stretch of the utterance over which coarticulation operates is of great interest, since it tells us how much of an utterance should normally cohere. Vowels with primary stress do not seem to be the primary source of coarticulatory spread (Magen, 1989), despite their initial appeal as the locus for cohesion. It is possible that lexical stress (even on the neologistic utterances used here) is governed by the constraints that apply to the segments, and that more discourse-oriented features such as focus are responsible for differences in coarticulatory span. Similarly, we should not rule out the possibility that individual differences in amount and organization of coarticulation will be detectable, perhaps contributing to the ease or difficulty some people have in making themselves understood. Knowing the bounds of such coarticulation would also further inform the improvement of speech synthesis by rule.

Without being explicit about the mechanisms of planning, it is still possible to find evidence for planning or its lack. In the present experiments, there is ample evidence of a great diminution of anticipatory coarticulation when a talker begins an utterance before knowing the influencing segment. Since normal coarticulatory effects were consistently found for other segments in those same conditions, we can be confident that the utterances were not simply unnatural throughout. Rather, talkers seem to plan for as much coarticulation as they can. If perseverative coarticulation is planned (as there is some evidence it is), then this planning can be accomplished before a syllable starts, in the course of a preceding vowel and stop closure. This is further evidence that talkers structure their utterances to include coarticulation, presumably because it serves a communicative function.

REFERENCES

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica, 38,* 9-20.

Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation: lip rounding. *Journal of the Acoustical Society of America, 71,* 449-454.

Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). New York: Academic Press.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,*219-252.

Browman, C. . P., & Goldstein, L. (in press). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* Cambridge: Cambridge University Press.

Canic, M. J., & Franks, I. M. (1989). Response preparation and latency in patterns of tapping movements. *Human Movement Science, 8,*123-139.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind, 11,* 220-242, 377-392, 524-538.

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22,* 129-159.

Dick, A. O. (1971). Processing time for naming and categorization of letters and numbers. *Perception & Psychophysics, 9,* 350-352.

Fant, G. (1973). Stops in CV-syllables. In G. Fant (Ed.),*Speech sounds and features* (pp. 110-139). Cambridge, MA: MIT Press.

Fant, G., Ishizaka, K., Lindqvist, J., & Sundberg, J. (1972). Subglottal formants. *Speech Transmission Laboratory Quarterly Progress and Status Report (Stockholm: Speech Transmissions Laboratory, KTH), 1,* 1-12.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception, *Perception & Psychophysics, 25,* 375-388.

Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 113-133.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica, 38,* 35-50.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception, *Perception & Psychophysics,36,* 359-368.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-139). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, C. A., & Vatikiotis-Bateson, E. (in preparation). Declination of amplitude in utterances showing declination of fundamental frequency.

Garcia-Colera, A., & Semjen, A. (1988). Distributed planning of movement sequences. *Journal of Motor Behavior, 30,*341-367.

Gay, T. J. (1977). Articulatory movements in VCV sequences. *Journal of the Acoustical Society of America, 62,* 183-193.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America, 76,* 18-26.

Hombert, J. M., Ohala, J., & Ewan, W. (1979). Phonetic explanation for the development of tones. *Language, 55,* 37-58.

Izdebski, K., & Shipp, T. (1978). Minimal reaction times for phonatory initiation. *Journal of Speech and Hearing Research, 21,* 638-651.

Keating, P. A. (1985). CV Phonology, experimental phonetics, and coarticulation. *UCLA Working Papers in Phonetics, 62,* 1-13.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology, 5,* 275-292.

Kent, R. & Minifie, F. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5,* 115-133.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code, *Psychological Review, 74,* 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII, Perception* (pp. 143-178). New York: Springer-Verlag.

Lisker, L. (1986) "Voicing" in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech, 29,*3-11.

Löfqvist, A., Baer, T., McGarr, N., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85,* 1314-1321.

Magen, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America, 75,* S41.

Magen, H. (1989). *An acoustic study of vowel-to-vowel coarticulation in English.* Unpublished doctoral dissertation, Yale University.

Manuel, S. Y. (1987). *Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages.* Unpublished doctoral dissertation, Yale University.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. Haskins Laboratories *Status Report on Speech Research, SR77/78,* 69-78.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244,* 522-523.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 473-488.

Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 72-86). Berlin: Springer.

Ohde, R. N. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America, 75,* 224-230.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151-168.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America, 76,* 1624-1635.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics, 15,* 299-312.

Recasens, D. (1989). *Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences.* Manuscript submitted for publication.

Repp, B. H., & Lin, H.-.B. (1987). Difference in second-formant transitions between aspirated and unaspirated stop consonants preceding [a], *Language and Speech, 30,* 115-129.

Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 129-144). Berlin: Springer.

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94,* 84-106.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1(4),* 333-382.

Semjen, A., & Garcia-Colera, A. (1986). Planning and timing of finger-tapping sequences with a stressed element. *Journal of Motor Behavior, 18,* 287-322.

Sharf, D. J. (1962). Duration of post-stress inter-vocalic stops and preceding vowels. *Language and Speech, 5,* 26-30.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 469-505). New York: Academic Press.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *The perception and production of fluent speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language, 65,* 81-106.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America, 82,* 847-863.

Watson, B. C., & Alfonso, P. J. (1983). Foreperiod and stuttering severity effects on acoustic laryngeal reaction time. *Journal of Fluency Disorders, 8,* 183-205.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49-64.

Wolf, C. G. (1978). Voicing cues in English final stops. *Journal of Phonetics, 6,* 299-309.

FOOTNOTE

Journal of Phonetics, 18, 3-35 (1990).

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102,177-193

The Perceptual Effects of Child-Adult Differences in Fricative-Vowel Coarticulation*

Susan Nittrouer and D. H. Whalen

Earlier work [Nittrouer, Studdert-Kennedy, and McGowan (1989)] demonstrated greater evidence of coarticulation in the fricative-vowel syllables of children than in those of adults when measured by anticipatory vowel effects on the resonant frequency of the fricative back cavity. In the present study, three experiments showed that this increased coarticulation led to improved vowel recognition from the fricative noise alone: Vowel identification by adult listeners was better overall for children's productions, and was successful earlier in the fricative noise. This enhanced vowel recognition for children's samples was obtained in spite of the fact that children's and adults' samples were randomized together, therefore indicating that listeners were able to normalize the vowel information within a fricative noise where there often was acoustic evidence of only one formant associated primarily with the vowel. Correct vowel judgments were found to be largely independent of fricative identification. However, when another coarticulatory effect, the lowering of the main spectral prominence of the fricative noise for /u/ versus /i/, was taken into account, vowel judgments were found to interact with fricative identification. The results show that listeners are sensitive to the greater coarticulation in children's fricative-vowel syllables, and that, in some circumstances, they do not need to make a correct identification of the most prominently specified phone in order to make a correct identification of a coarticulated one.

INTRODUCTION

Coarticulation, the temporal overlap of the production and acoustic realization of phonemes, helps explain the difficulty in finding acoustic invariance in the speech signal and the inappropriateness of taking acoustic segments to correspond to linguistic segments (Liberman et al., 1967; Studdert-Kennedy, 1976, 1985; Liberman & Mattingly, 1985). Not only are these effects obvious and far-reaching in the acoustic record, but there is also ample evidence that listeners are sensitive to the information coarticulation provides. Of particular interest for the present work is the vowel information present in neighboring fricative noises (e.g., LaRiviere et al., 1975; Soli, 1981; Yeni-Komshian & Soli, 1981; Whalen, 1983). For adult listeners, there is often enough vowel information in these fricative noises for reasonably accurate identification of the vowel from the fricative noise alone.

While coarticulation, and its perceptual effect, has been well-studied for adults, only recently has it been examined developmentally (Flege, 1988; Nittrouer, 1985; Nittrouer & Studdert-Kennedy, 1987; Nittrouer et al., 1989; Sereno et al., 1987). The examination by Nittrouer et al. (1989) of fricative-vowel syllable production has revealed several interesting similarities and differences between children and adults. One coarticulatory effect, the co-occurrence of liprounding for the vowel with a preceding fricative gesture, was found to be the same for both groups of speakers. Positioning of the tongue during fricative production, however, was found to be more vowel-dependent for the children than for the adults. As

---

This work was completed while the first author was a postdoctoral fellow at Haskins Laboratories, supported by NIH Grant NS-07237 to Haskins Laboratories. It was also supported by NIH Grant HD-01994 to Haskins Laboratories. We thank Carol Fowler, John Mott, Donna Neff, Michael Studdert-Kennedy, and two anonymous reviewers for their helpful comments.

for the fricatives themselves, there was less differentiation of the fricatives /ʃ/ and /s/ for the children's productions than for the adults'. The first of these findings (i.e., that the magnitude of liprounding was similar for child and adult speakers) was reported also by Sereno et al. (1987). However, their results failed to reveal any difference between children and adults in the magnitude of lingual coarticulation. The reason for this discrepancy in findings may be that Sereno et al. (1987) limited the frequency region in which they looked for and measured the spectral consequence of lingual coarticulation, whereas Nittrouer et al. did not. Sereno et al. did not investigate the precision of children's fricative production relative to that of adults.

Two acoustic measurements led to the three major conclusions of Nittrouer et al. (1989). First, centroid measurements demonstrated greater spectral similarity for children's samples of /ʃ/ and /s/ noises than for adults'. The centroid, or first spectral moment, is a frequency mean of the whole spectrum, weighted by amplitude, and seems to be a good index of the distinctiveness of fricative production (Forrest et al., 1988; Nittrouer et al., 1989). Of particular concern to the present study, centroids appear to reflect reliably the frequency of the fricative spectral prominence, which is the formant (or formants) affiliated with the constriction and the cavity anterior to it. Nittrouer et al. found children's centroids for /s/ and /ʃ/ to be intermediate between adults' /s/ and /ʃ/ centroids, with both relatively closer to adults' /s/ values than to their /ʃ/ values.

Centroid measurements also showed that children and adults liprounded in anticipation of upcoming rounded vowels to the same extent during fricative production. Liprounding lengthens the cavity anterior to the constriction, lowering centroid values. Similar shifts in these values were found for children and adults for syllables containing the vowel /u/ as compared to those containing the vowel /i/, and these shifts were greater for /s/ than for /ʃ/ noises across all age groups.

The finding of greater coarticulation of lingual vowel gestures for children's productions than for adults' was supported by measurements of fricative second formants (F2s). Fricative F2 frequencies varied as a function of the upcoming vowel to a greater extent for children's samples than for adults'. Across speaker age, however, there was overlap in fricative F2 frequencies between the two vowel categories: Children's samples for both /i/ and /u/ fell within or above the adult /i/ range. Because fricative F2 is affiliated with the cavity posterior to the fricative constriction it primarily indexes the amount of anticipatory lingual posturing. Relatedly, McGowan and Nittrouer (1988) showed that the amplitude of fricative F2, relative to the higher-frequency components, was greater for children's samples. In fact, it was difficult to identify an F2 resonance at all in adults' samples until 30 ms before vowel onset (cf. Soli, 1981). Therefore, it was not possible to determine how much earlier, if at all, this difference in lingual coarticulation extended back into the fricative.

It is conceivable that the two differences between children's and adults' FV-syllable production described by Nittrouer et al. (1989) were related, especially because both involve tongue gestures. That is, either the greater anticipatory coarticulation of the vowel tongue-body gesture may have prohibited proper tongue-tip placement for the fricative, or the less-precise fricative gesture may have permitted greater assimilation of the vowel gesture. In either case, high negative correlations would be expected between measures of the magnitude of vowel coarticulation (i.e., fricative F2s) and those of the precision of fricative production (i.e., centroids). However, Nittrouer et al. failed to find significant correlations between these measures. Thus, the precision of fricative production and the magnitude of coarticulation were not related in individual speakers.

In the present study, we extended the work of Nittrouer et al. (1989) to answer three questions regarding the perception of coarticulatory effects in the FV syllables studied by them:

1) Do the enhanced coarticulatory effects of children's FV syllables provide usable perceptual information, or are they just noise?
2) If children's enhanced coarticulation is perceptible, do the perceptual tasks indicate that the age-related differences are greater in temporal extent or only in magnitude?
3) Are there interactions between fricative and vowel judgments when those judgments are based on the fricative noise alone?

The first question is of interest because the acoustic measurements of Nittrouer et al. (1989) not only indicated that coarticulation was greater for children's FV syllables, but also that the acoustic difference between the two fricatives studied (/s/ and /ʃ/) was less. Sereno et al. (1987) presented adults' and children's fricative noises in isolation, and found that vowel judgments by adult listeners were more accurate for the adult talkers. Therefore, it may be that any vowel information provided by the fricative F2 is uninterpretable without correct identification of the fricative. If this is the case we would expect identification of both the fricative and the vowel for children's fricative noises to be impaired. Conversely, if the greater acoustic evidence of vowel information in the fricative noise is in fact perceivable as vowel information, then vowel identification for children's samples should be better than for adults'. As noted above, decisions based on the F2 resonance will require a correct identification of the speaker as a child or as an adult because children's /u/ values are as high or higher than adults' /i/ values.

The second question has already received a partial answer in Nittrouer et al. (1989). They had difficulty even finding an F2 in some of the adult's productions, but found F2s, and vowel-related frequency effects on these values, quite early in the fricative noises of children's samples. McGowan and Nittrouer (1988) also found, for a subset of the stimuli used in the present experiments, that the intensity differences between fricative F2s in children's and adults' productions decreases from the beginning of the noise to vowel onset. Taken together, these results suggest that, at least for lingual gestures, children coarticulate more from the very onset of the syllable, and that age-related differences in production are less just before release of the fricative constriction. The effects of liprounding, however, have not been studied throughout the fricative noise, so it may be that there are differences at the beginning of the noise for this effect as well. Specifically, children may have coarticulated liprounding to the same degree as adults, but starting earlier.

The final question concerns sources of information used for making fricative and vowel judgments, and how these judgments interact. Two acoustic sources of information were of interest here: fricative F2 and the higher spectral prominence. Because an earlier study (Harris, 1958) demonstrated that judgments for /ʃ/ and /s/ are not strongly dependent on F2 frequency just after vowel onset, it is reasonable to think that these fricative judgments would not be strongly dependent on F2 frequency within the fricative noise itself. However, the vowel also affects the main spectral prominence of the fricative noise (Mann & Repp, 1980; Whalen, 1981), and fricative judgments of the noise can affect how the vowel is perceived (Whalen, in press). Two of the phones (/u/ and /ʃ/) are associated with lower-frequency noise, while the other two (/i/ and /s/) are associated with higher-frequency noise. When all four combinations of fricative and vowel are considered, we find that two combinations (/ʃi/ and /su/) allow for a trade between fricative and vowel judgments: Noises lacking extreme frequency values can be heard either as an /i/-colored /ʃ/ or as an /u/-colored /s/. Such a trade has been found for synthetic speech (Whalen, in press), and so we will be looking for it in the responses to our truncated natural speech.

Each of these questions was studied by having adult listeners make phonetic judgments of syllable components based, often, on just a portion of the fricative noise. In general, we expected to confirm that children differ in their organization of speech gestures (Nittrouer et al., 1989), but do so in a way that nonetheless leads to correct recovery of phonetic information by adult listeners. Our results will thus substantiate or refute the age-related acoustic and articulatory differences reported by Nittrouer et al. and by McGowan and Nittrouer (1988). They will begin also to tease apart the processes of recognizing coarticulated vowels and consonants.

General Method

A. Stimuli

All stimuli for the present set of experiments came from speech samples collected for Nittrouer et al. (1989), and additional details can be found there. Samples consisted of the disyllables /ʃiʃi/, /ʃuʃu/, /sisi/, and /susu/, spoken by eight adults (four female, four male) 20 or 21 years old, and eight children at each of the ages 3, 4, 5, and 7 years. No greater than a 3/5 ratio existed between the numbers of boys and girls in any of the children's age groups. All speakers had hearing thresholds better than 20 dB HL at each of the frequencies 500 Hz, 1 kHz, 2 kHz, and 4 kHz, and had normal articulation for their ages, as judged independently by two speech pathologists. The samples were spoken in response to pictures of a girl, referred to by the pronoun "she," of a boy pointing and saying "see," of a shoe, and of a girl named "Sue." All pictures had been used with these subjects in an earlier perception experiment, and so were familiar to them. Although a short practice session was provided in which speakers imitated the experimenter's model of producing the syllables at a constant rate with equal stress on each syllable, recorded samples were not imitated, but were instead spoken in response to the pictures. This method was chosen to ensure samples as close as possible to each speaker's normal production. Samples were collected in randomized groups of four, and ten samples of each syllable were obtained. Of these, the first eight samples with no extraneous noise were digitized. Thus, 32 samples from each speaker were available for use in the present study.

A Uher model 4200 portable tape recorder with an Electrovoice model 635A microphone was used to collect the samples. This system had a flat frequency response out to 10 kHz. The eight samples of each utterance per speaker were digitized onto a VAX 11/780 computer at a sampling rate of 20 kHz with low-pass filtering below 9.6 kHz. Using the Haskins Laboratories Waveform Editing and Display program (WENDY), the onsets of both vowels in all tokens were identified as the points at which periodicity first appeared in the waveform.

WENDY was also used to extract the short speech slices to be used in the present experiments. Which tokens were used in each experiment depended on the purpose of the individual experiment. However, slices for all three experiments were extracted from the first syllable only, and were recorded on tape. No carrier phrase was used. The interstimulus interval (ISI) was always 2.5 s, with a 5-sec pause after every 10 tokens.

B. Listeners

All listeners in the present experiments were native speakers of American English between the ages of 20 and 40 years with no history of speech or hearing disorders, and were naive to the purpose of the experiment in which they participated.

C. Procedure

Listeners sat in a quiet room and listened to stimuli at a comfortable level over TDH-39 headphones. Instructions were given verbally prior to the start of the experiment, with the response alternatives indicated to the subjects. All responses were written.

EXPERIMENT 1

This experiment was designed primarily to answer the first question posed in the Introduction: Did the enhanced coarticulatory effects found by Nittrouer et al. (1989) for children's FV syllables provide usable information? It also provided preliminary data regarding the other two questions.

A. Method

Two speakers (one male and one female) in each age group from the Nittrouer et al. (1989) study were chosen at random. Two tokens of each utterance from each of these speakers were selected randomly from the set of eight, and five slices were extracted from the first syllable of each:

1. the first half of the fricative (1/2F)
2. the first three quarters of the fricative (3/4F)
3. the entire fricative (allF)
4. the fricative, plus the first quarter of the vowel (F+1/4V)
5. the entire syllable (allFV).[1]

These forty slices per speaker (4 utterances × 2 tokens × 5 slices) were recorded six times each in randomized blocks of 400 (40 slices × 10 speakers), giving a total of 2400 items. In order to determine whether or not fricative and vowel judgments interact to any extent, the five listeners participating in this experiment were asked to identify the syllable from which they thought the slice was extracted (i.e., "see," "she," "sue," or "shoe"), not just the upcoming vowel. If listeners had not been required to make fricative judgments, or if fricative identity had been given to them, it would not have been possible to determine whether or not correct vowel judgments were obtained in the absence of correct fricative judgments. Also, it would not have been possible to substantiate or refute the claim of Nittrouer et al. (1989) that children were less precise than adults in fricative production.

B. Results

Analyses of variance were performed on the data from Experiment I using speaker age, consonant, vowel, and slice as main effects, and Adults vs. Children as a planned comparison.[2] For the analyses reported here, data were collapsed over repetitions of the same token, tokens of the same utterance, and speakers within an age group so that the degrees of freedom would not be artificially inflated. Scores were submitted to an arcsine transform because fricative judgments were nearly 100% correct.

Whole-syllable identification. Figure 1 shows the percentage of correct whole-syllable responses to each slice for each age group. Responses were significantly above chance, as indicated by filled symbols, for all slices for all age groups, with scores improving as slices increased in length. Scores were not statistically different for adult speakers than for child speakers for slices consisting only of fricative noise: The mean number of correct responses to these three slices was 61% for adult samples and ranged between 52% and 69% for samples from the children's groups. This nonsignificant finding suggests that the enhanced fricative information of the adults' samples over the children's samples and the enhanced anticipatory vowel information of the children's samples over the adults' contributed about equally to whole-syllable judgments made from the fricative noise alone. Statistically significant results were obtained for the main effects of fricative (F(1,240) = 58.21, $p < .01$), vowel (F(1,240) = 7.31, $p < .01$), and slice (F(2,240) = 72.00, $p < .01$), indicating that syllables were more accurately recognized when the fricative was an /s/ rather than an /ʃ/ (66% vs. 55%), when the vowel was an /i/ rather than an /u/ (62% vs. 58%), and when more of the fricative noise was presented (70% for allF vs. 53% for 1/2F).

For the two slices containing some portion of the vocalic segment (F+1/4V and allFV), responses to adults' samples were more accurate than responses to children's samples: The mean number of correct syllable judgments for these two slices was 95% for adults' samples and ranged between 79% and 91% for samples from children's groups. An analysis of variance of whole-syllable scores for these two longest slices showed a significant main effect of age (F(4,160) = 24.34, $p < .01$) and a significant result for the planned comparison of Adults vs. Children (F(1,160) = 58.22, $p < .01$). In addition, significant main effects were found for fricative (F(1,160) = 44.66, $p < .01$), vowel (F(1,160) = 7.50, $p < .01$), and slice (F(1,160) = 52.29, $p < .01$). Again, more accurate syllable judgments were obtained for tokens with the fricative /s/ (91% vs. 84% for /ʃ/) and with the vowel /i/ (88% vs. 87% for /u/), and for longer slices (91% for allFV vs. 84% for F+1/4V).

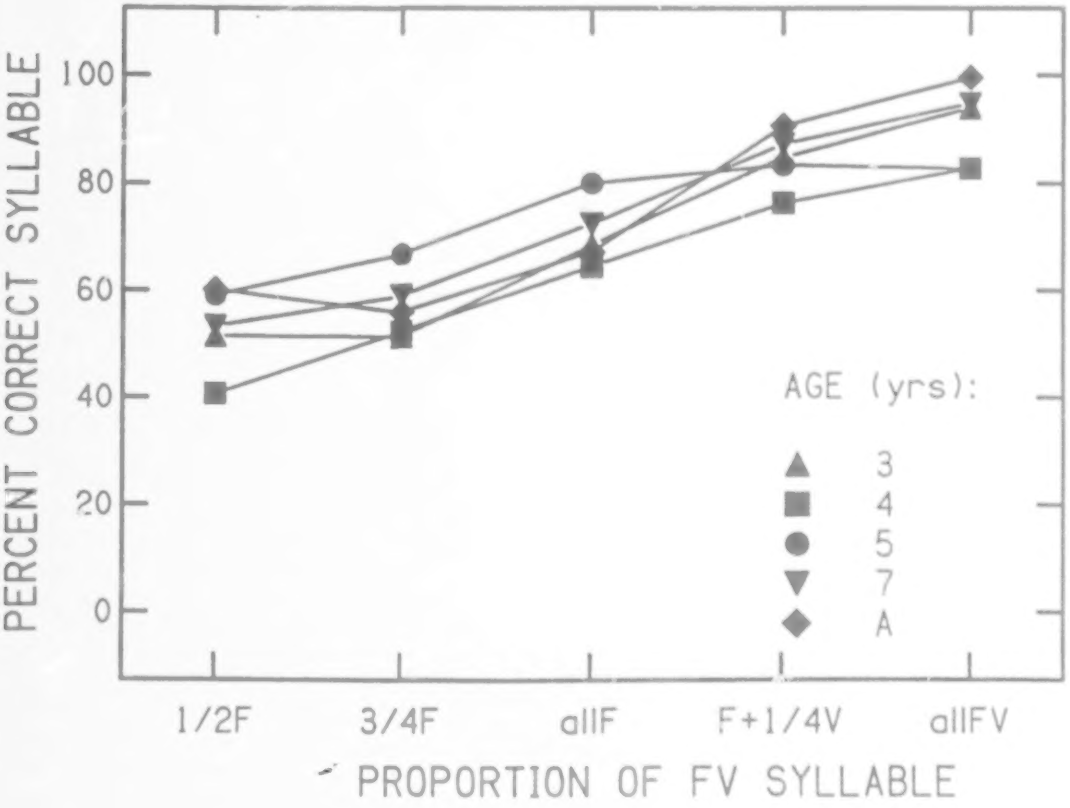


Figure 1. Mean percent correct whole-syllable identification scores from Experiment 1 for slices extracted from FV syllables spoken by children of 4 ages, and by adults. Filled symbols indicate scores that were significantly above chance ($p < .05$).

Fricative identification. Figure 2 displays the percentages of correct fricative responses (i.e., whole syllable correct plus fricative only correct). Again, all scores were clearly well above chance, but here it can be seen that responses to adults' slices were always superior to those for children's slices: Averaging across all slices, the mean number of correct fricative judgments was 99% for adults' samples, and ranged between 81% and 94% for samples from children's groups. An analysis of variance performed on fricative scores for all five syllable slices demonstrated that the main effect of age (F(4,400) = 64.32, *p* <.01) and the planned comparison of Adults vs. Children (F(1,400) = 182.01, *p* <.01) were both significant. In addition, the main effects of fricative (F(1,400) = 68.57, *p* <.01), vowel (F(1,400) = 32.25, *p* <.01), and slice (F(4,400) = 18.18, *p* <.01) were all found to be significant, reflecting more accurate fricative judgments for /s/ (95% vs. 87% for /ʃ/), for the /u/ context (94% vs. 89% for the /i/ context), and for longer slices (93% for allFV vs. 86% for 1/2F).

Vowel identification. In Figure 3, the percentages of correct vowel responses (i.e., whole syllable correct plus vowel only correct) are shown, all of which are significantly above chance. It can be seen that, for slices consisting only of fricative noise, responses to children's tokens were generally more accurate than responses to adults' tokens. For slices consisting of 3/4 of the fricative noise, only responses to 3-year-olds' samples were as poor as those to adults' samples. For slices consisting of the whole fricative, responses were better to samples from all children's groups than to those from adults: Percentage of vowels correctly recognized was 67% for adults and 78% for children, averaged over all four age groups. An analysis of variance performed on vowel identification scores for the three slices consisting only of fricative noise demonstrated significant results for the main effect of age (F(4,240) = 9.25, *p* <.01) and the planned comparison of Adults vs. Children (F(1,240) = 13.43, *p* <.01). Also found to be significant were the main effects of fricative (F(1,240) = 8.04, *p* <.01), vowel (F(1,240) = 75.12, *p* <.01), and slice (F(2,240) = 79.53, *p* <.01). These significant effects indicate that listeners more often identified vowel context correctly for /s/ than for /ʃ/ noises (69% vs. 65%), for the /i/ than for the /u/ context (72% vs. 61%), and for longer than for shorter slices (76% correct for allF vs. 61% for 1/2F).

For the two slices which contained some part of the vocalic segment, the greater accuracy of vowel judgments for children's samples disappears. Obviously these slices provide much more information about vowel identity than the three slices consisting only of fricative noise. Puzzlingly, identification of the 5-year-olds' vowels did not improve with the addition of the vocalic segment.

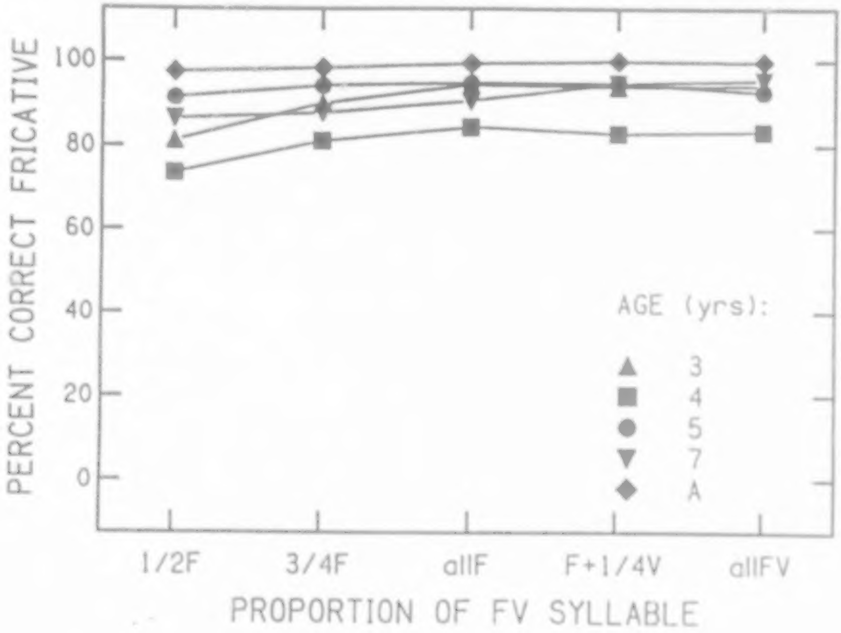


Figure 2. Mean percent correct fricative identification scores from Experiment 1. (See Figure 1 legend for details.)

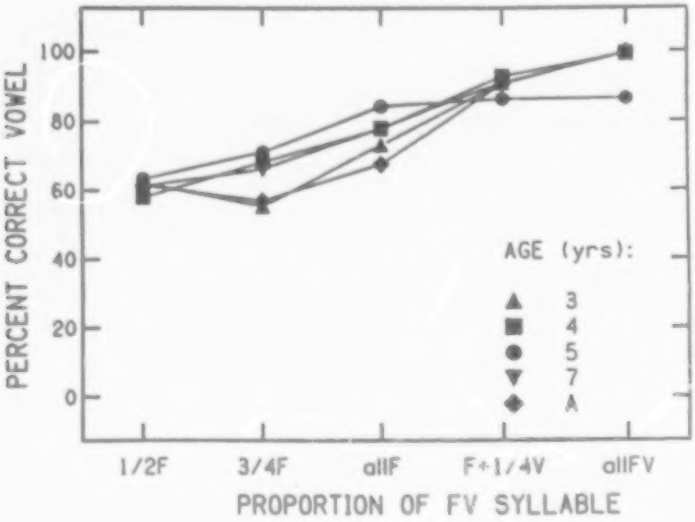


Figure 3. **Mean percent correct vowel identification scores from Experiment 1. (See Figure 1 legend for details.)**

C. Discussion

This first experiment showed that adult listeners were indeed able to interpret correctly the vowel information in the fricative noises of children's productions. This was true despite the fact that identification of the children's fricatives was consistently worse than that of the adults'. Identification of the vowel was generally better for children's productions when only the fricative noise was given, but was of essentially the same accuracy for adults' and children's samples when some or all of the vocalic segment was present. This confirms the acoustic results (Nittrouer et al., 1989) that indicated greater anticipatory coarticulation in children's FV syllables than in adults'. In addition, the lower identification scores for children's fricatives support the claim of less distinct fricative production in children's FV syllables. The somewhat greater accuracy for recognition of /s/ than of /ʃ/ may reflect the fact that although children's centroids for both /s/ and /ʃ/ were intermediate between adults' /s/ and /ʃ/ values, children's /ʃ/ centroids showed disproportionately higher values. However, it would be inappropriate to place too much importance on these significant fricative effects because, as will be seen, they were not found in Experiment 2.

The basis for the correct vowel judgments must be inferred, as we do not have any direct measure of it. For the adult productions, especially for judgments on the first half of the fricative noise, the effect of the vowel on the main spectral prominence of the fricative noise is the likeliest source. Adults show very little, if any, trace of F2 that early in the noise, yet vowel judgments were above chance even for that slice. The finding of slightly more accurate vowel recognition for /s/ noises than for /ʃ/ noises also suggests that listeners could have used the vowel-related shifts in the fricative spectral prominence: these shifts were greater for /s/ than for /ʃ/ (Nittrouer et al., 1989). Of course, use of these shifts in the fricative spectral prominence to correctly identify vowel context required correct fricative identification, which was not always the case for children's samples. For the children's productions, it seems more likely that F2 was used throughout, given that the fricative was often misidentified. Even if the fricative is misidentified, F2 provides fairly accurate information about vowel context. Thus, the finding that listeners could correctly identify the vowel at all for adults' samples coupled with the difference in identification accuracy between adults' and children's productions indicates that both sources of vowel information were used.

In interpreting the F2 resonance in the fricative noise, our listeners must have been doing some type of normalization. This is because the absolute values for the F2 in children's fricative noises are consistently higher than those of adult's, regardless of vowel category. Unless vocal tract size was taken into account, an F2 from a child's production would always be identified as an /i/ resonance. Thus, even when only a fricative noise is present, listeners correctly account for vocal tract size, despite the absence of a clear formant structure.

It could be argued that the finding of a significant main effect of vowel identity on vowel judgments suggests this very bias of perceiving all children's samples as coming from an /i/ context. Since there were four times as many children's as adults' samples in this experiment, it may actually account for some portion of this effect. However, if it were completely responsible for the enhanced recognition of /i/ over /u/, then vowel recognition for /u/ would have been significantly below chance, which it was not. More likely, because fricative F2 is higher in frequency preceding /i/ than preceding /u/, it is closer to the main spectral prominence of the fricative, and so, is also greater in relative amplitude. Therefore, fricative F2 preceding /i/ is more easily accessible to listeners.

EXPERIMENT 2

Having answered the first question we posed in the Introduction with Experiment 1, we designed Experiment 2 to investigate more carefully the next two questions, concerning the time course of children's enhanced coarticulation and the interaction of fricative and vowel judgments. Although some indication of temporal effects had been found in the first experiment, overall identification scores were too high to see specifically what those effects were. The fact that overall identification scores were so high also prohibited close inspection of the interaction between fricative and vowel judgments. Moreover, the fact that the duration of slices varied as a function of the length of the fricative noise could have been a confounding factor. We overcame these difficulties in the second experiment by presenting very brief slices of the same absolute duration from different locations within the fricative noise.

In addition, we wanted to examine more closely the relation between the perceptual results and earlier acoustic findings in this second experiment. Although we had found a relation in the first experiment between speaker age and identification scores, both for fricative and for vowel, we were not able to discern exactly how these perceptual differences were related to the age-dependent acoustic (and therefore, articulatory) differences reported by Nittrouer et al. (1989) and by McGowan and Nittrouer (1988). One reason for this shortcoming was the selection of speakers was random, and, as happens with random selection, some of the age groups were represented by less-adept speakers. In the present experiment, then, we selected children on the basis of the acoustic measure of fricative differentiation, which Experiment 1 showed to be a possible factor in the perception of the fricative, since children's fricatives were less accurately perceived.

A. Method

Three groups of speakers were chosen from those participating in Nittrouer et al. (1989). One group consisted of four adults (two male and two female), selected randomly. A second group consisted of four children, one child of each of the ages sampled (3, 4, 5, and 7 years) who demonstrated less than a 1-kHz difference in mean centroids between /ʃ/ and /s/. Because this lack of spectral differentiation between fricatives was typical of children's samples, this was labelled the "child-like" group. The third speaker group consisted of two children (4- and 5-years old) who demonstrated greater than a 1-kHz difference between/ʃ/ and /s/ centroids.[3] Because this pattern was more similar to adult production, this was labelled the "adult-like" group. The two groups of children did not differ from each other in either extent to which they coproduced the following vowel with the fricative or the amplitude of fricative F2 relative to the higher formants. Thus, samples from the adult-like children were similar to those from the adults on the extent to which the fricative spectral prominence provided information about fricative identity, but were similar to samples from the child-like children on the extent to which F2 provided information about vowel identity.

All eight tokens of each utterance from each speaker were used. Slices 60 ms in duration were extracted beginning just before vowel onset, and then working backwards into the fricative in 40 ms steps until the beginning was reached. Between two and seven slices were extracted from each fricative (i.e., fricatives were between 100 ms and 300 ms in duration). Children demonstrated more variability in fricative duration than adults, who consistently produced fricatives between 150 ms and 220 ms long. Extracted slices were randomized, and a test tape consisting of 1273 slices of fricative noise was made. Eight new listeners participated in this experiment, and their task was to write the syllable, both fricative and vowel, from which they thought the slice was extracted.

B. Results

Scores were collapsed over tokens of the same slice (i.e., taken from the same location) within the same utterance, and over speakers within a group as a way of restricting the number of degrees of freedom. These values were entered into analyses of variance with the main effects of speaker group, fricative, vowel, and slice. In addition, orthogonal planned comparisons were performed for Adults vs. Children and Adult-like vs. Child-like Children. Arcsine transforms were not used in this experiment because identification scores were not close to either 100% or 0%.

While all slices extracted from each token were presented in the perception experiment, only responses to the five slices closest to vowel onset (-220 to -60 ms from vowel onset) were included in the analyses. Informal inspection for the few slices earlier than those revealed little difference between them and the -220 ms slice. Because their inclusion would have greatly increased the number of empty cells in the analysis, they were, instead, excluded altogether.

Whole-syllable identification. Figure 4 shows scores for whole syllable identification. All were significantly above chance, as indicated by the filled symbols, but there were differences among speaker groups. Although listeners were able to identify the syllable from which slices were extracted with equal accuracy for samples from the adult-like speakers and the adults (45% correct across all slices for each group), they were less accurate in syllable identification for slices from the child-like speakers (36% correct). This difference shows up in the statistical analysis as a significant main effect for speaker group (F(2,420) = 14.04, *p* <.01) and as a significant result for the comparison between Child-like and Adult-like Children (F(1,420) = 20.93, *p* <.01). It is not clear why a difference in syllable identification was found in this experiment for fricative noises from adults and from one group of children when no difference was found for adult and child speakers in Experiment 1. It could be that enough children in the previous experiment differentiated between /ʃ/ and /s/ in an adult-like manner that they could be considered adult-like speakers. However, the more likely explanation for the discrepancy in results between the two experiments is that the use of very brief slices in the present experiment made the task more difficult, and so, more sensitive to certain listener effects. This explanation is supported by the fact that, for the briefest slices in Experiment 1, syllable identification scores appear to be more accurate for slices from adults' samples than for those from any of the children's groups (see the 1/2F data of Figure 1). The only other effect to demonstrate statistical significance for syllable identification scores in Experiment 2 was the main effect of vowel (F(1,420) = 13.76, *p* <.01), reflecting the trend also found in Experiment 1 of more accurate syllable recognition in the /i/ context than in the /u/ context (45% vs. 39%).

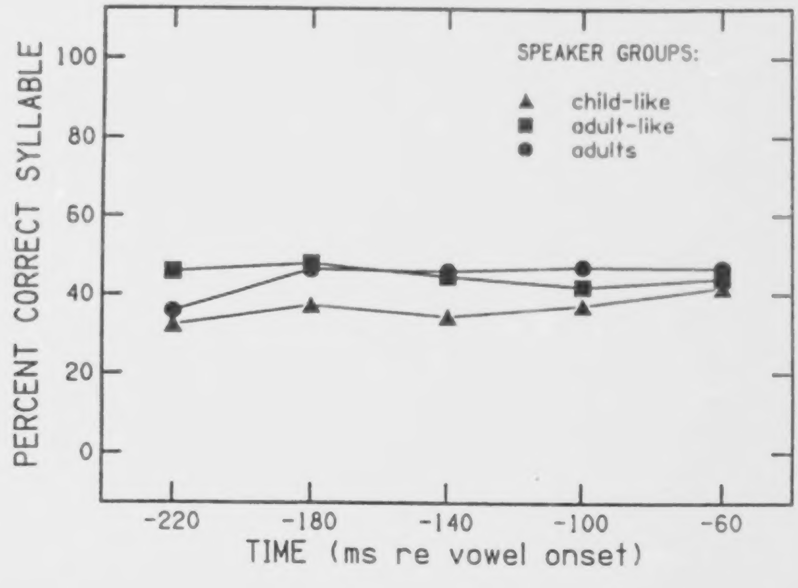


Figure 4. Mean percent correct syllable identification scores from Experiment 2 for slices of fricative noise extracted from FV syllables spoken by children and adults. Filled symbols indicate scores that were significantly above chance (*p* <.05).

Fricative identification. Figure 5 shows scores for fricative identification. All were significantly above chance, as indicated by the filled symbols. Listeners had the easiest time recognizing the fricative for adults, followed by adult-like children, and lastly by child-like children, and this effect of speaker group was statistically significant (F(2,420) = 81.59, p <.01), as were the comparisons of Adults vs. Children (F(1,420) = 137.75, p <.01), and Adult-like vs. Child-like Children (F(1,420) = 25.44, p <.01). Furthermore, overall fricative judgments decreased in accuracy as vowel onset was approached, as indicated by a significant main effect of slice (F(4,420) = 6.48, p <.01). Neither the main effect of fricative nor that of vowel was found to be significant.

Vowel identification. Figure 6 shows scores for vowel identification. Only scores for those slices closest to vowel onset were significantly above chance for all speaker groups. Identification scores at this location were 57% for adults, 67% for adult-like children, and 65% for child-like children. So, the 10% difference in vowel identification scores found in Experiment 1 is replicated here. In addition, scores for the child-like group for slices at -180 ms and -100 ms before vowel onset were significantly above chance, as were scores for slices between -180 ms and -100 ms of vowel onset for the adult-like group. This difference in vowel identification scores for speaker groups was statistically significant (F(2,420) = 7.91, p <.01), as was the comparison of Adults vs. Children (F(1,420) = 9.69, p <.01). The main effect of how close to vowel onset the slice was located was also significant (F(4,420) = 7.12, p <.01). Averaging across speaker groups, 50% of all vowel judgments were accurate for slices starting at -220 ms from vowel onset, while 63% of these judgments were accurate for slices starting at -60 ms. As found in Experiment 1, vowel judgments for /i/ were more accurate than those for /u/ (60% vs. 51%), and this difference was statistically significant (F(1,420) = 28.20, p <.01). No difference was found between /s/ and /ʃ/ for accuracy of vowel judgments.

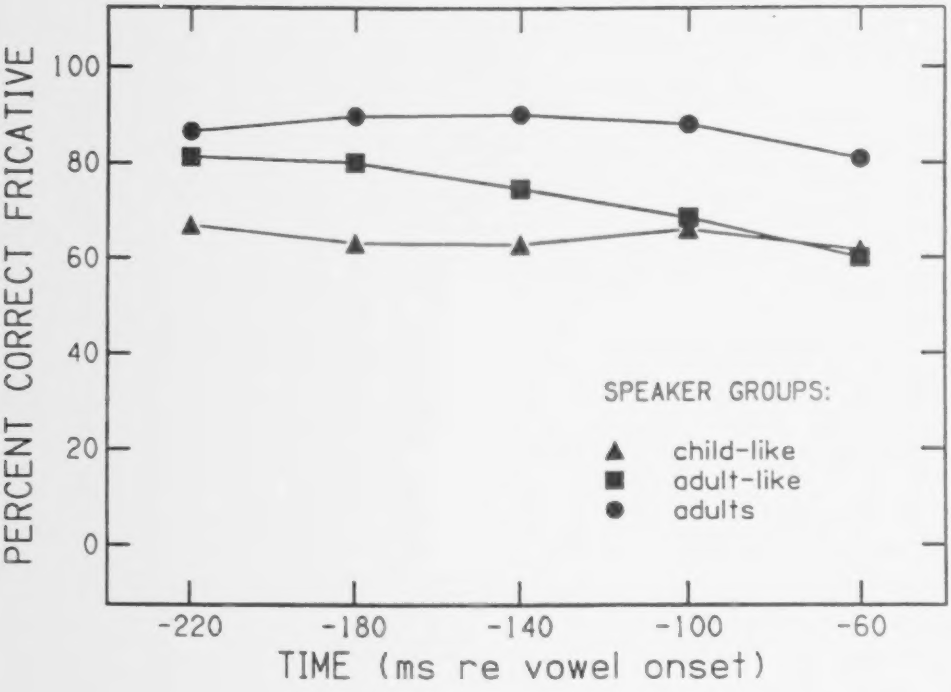


Figure 5. Mean percent correct fricative identification scores from Experiment 2. (See Figure 4 legend for details.)

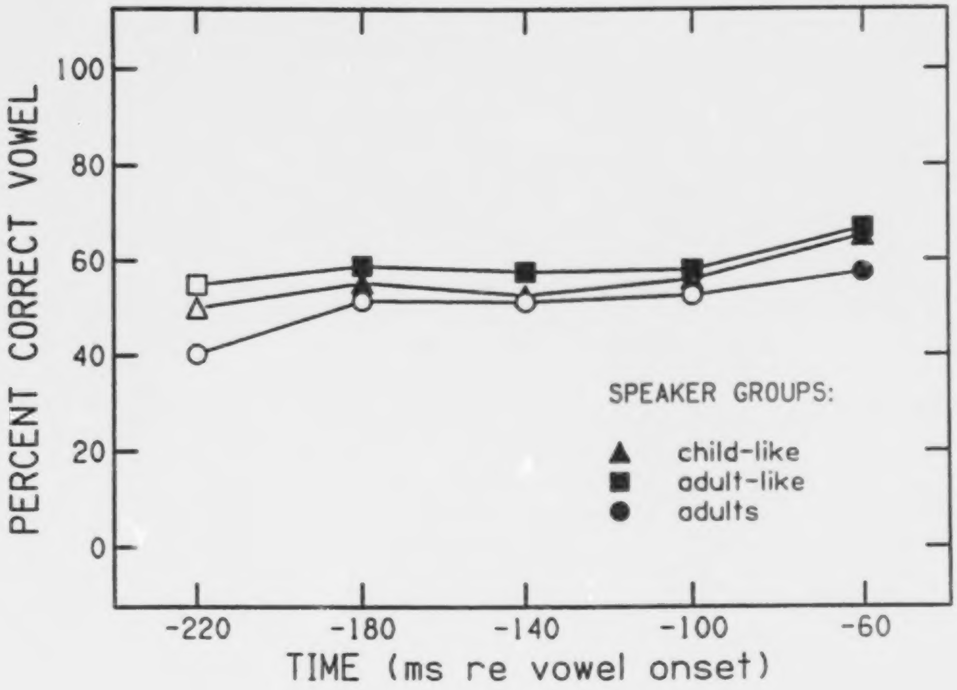


Figure 6. **Mean percent correct vowel identification scores from Experiment 2. (See Figure 4 legend for details.)**

C. Discussion

In general, the results of Experiment 2 support the interpretations of Experiment 1. Fricative identity was easier to recover from adults' productions than from children's, but vowel context was recovered both more reliably and earlier from children's. The difference in fricative identification between the adult-like and the child-like speakers confirms that the acoustic difference found previously (Nittrouer et al., 1989) has a straightforward effect on the perception of the fricatives. For vowel identification, samples from the adult-like children were not adult-like, but rather demonstrated more accurate judgments beginning earlier in the fricative noise, as did samples from the child-like group. This finding makes it even more likely that the greater accuracy for vowel judgments was due to the F2 resonance, and not to the fricative spectral prominence. On the acoustic measures of fricative F2, the two groups of children were similar.

From the results of Experiment 2, it seems that fricative information decreases over time: There was a general decline in fricative perceptibility as vowel onset was approached. The vowel information shows almost the reverse pattern. While the earliest time slice showed no significant vowel identification for any age group, the next three time slices showed a slight trend for increasing accuracy, and, on the last time slice, vowel identification scores improved for all three groups (to being the only slice from the adult speakers to elicit above-chance performance). These temporal effects of decreasing accuracy for fricative judgments with increasing accuracy for vowel judgments seem to be due to changes in posturing of the tongue, both in preparation for vowel production, and in termination of fricative production. Specifically, the tongue body takes on more of the position for the vowel, and the tongue-tip constriction is widened. This lowers the relative amplitude of the fricative spectral prominence while increasing the relative amplitude of the F2 resonance.

As mentioned previously, correct identification of the vowel for children's productions was found in the face of misidentifications of the fricative. For the most part, then, the two decisions were independent of each other. Yet the effect of the vowel on the main spectral prominence of the noise is such that an interaction between the decisions might be expected. The main

prominence of the noise is lower in frequency before /u/ than before /i/, and /ʃ/ noises are lower in frequency than /s/ noises. If, then, fricative and anticipatory vowel judgments made from the acoustic slices presented here were to any extent interdependent, the following pattern of predictions could be made: When the fricatives of /ʃu/ or /si/ were misidentified, the probability of correct vowel judgments would have increased. When the fricatives of /ʃi/ or /su/ were misidentified, the probability of correct vowel judgments would have decreased. That is, a low noise from /ʃ/ heard as /s/ should have led to /u/ vowel judgments, while a high noise from /s/ heard as /ʃ/ should have led to /i/ vowel judgments. Whether these judgments were correct or not depended, of course, on what the original context actually was.

Figures 7 through 9 display the percent correct vowel identification for all instances of incorrect fricative identification. Responses to each syllable type are shown separately. Filled symbols indicate that performance was significantly greater or lesser than chance. Adult speakers (Figure 7) and adult-like speakers (Figure 8) show good agreement with the predictions. Of the 40 possible comparisons, 31 are significantly different from chance, and 28 of the 31 are in the predicted direction: Scores for /ʃi/ and /su/ are significantly poorer than chance, while scores for /ʃu/ and /si/ are significantly better. Thus, there is substantial evidence for an interaction between the vowel and the fricative judgments. Even with the child-like speakers (Figure 9), nine of 14 significant comparisons are in the predicted direction, although not as strongly.

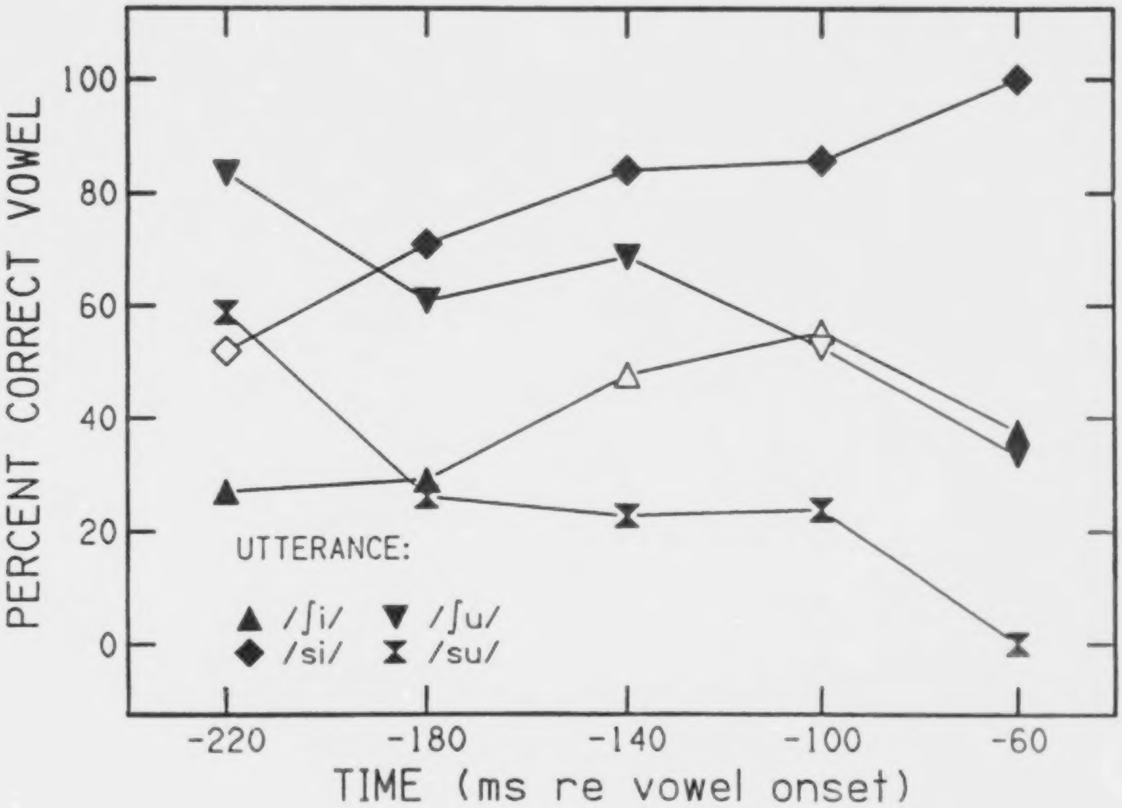


Figure 7. Mean percent correct vowel identification when the fricative was misidentified for samples from adult speakers. Data are from Experiment 2, and scores significantly different from chance are indicated by filled symbols.

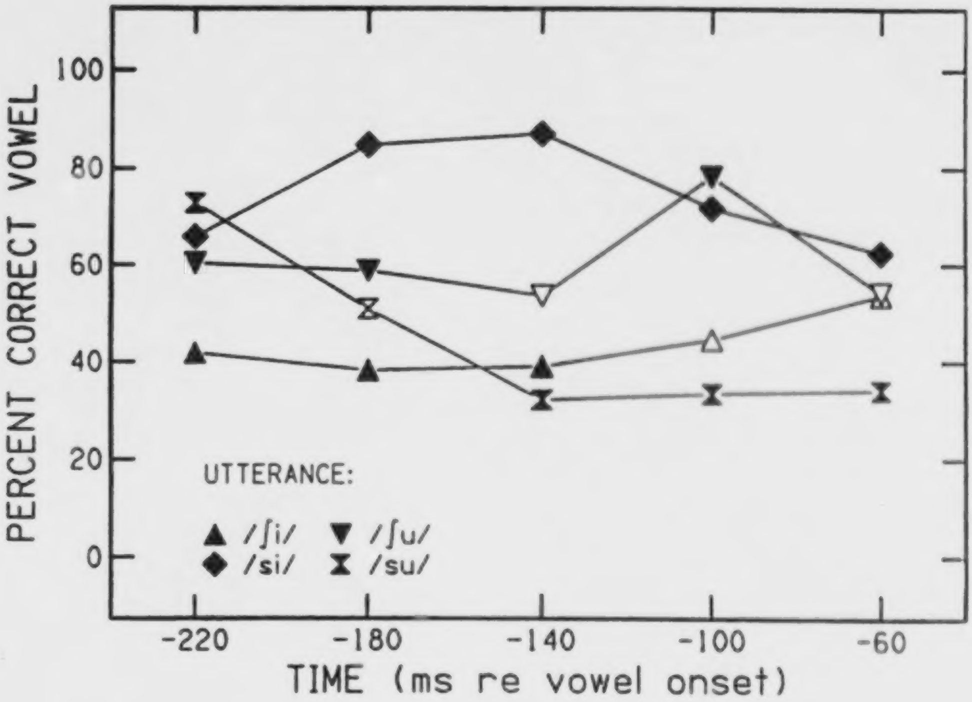


Figure 8. **Mean percent correct vowel identification when the fricative was misidentified for samples from adult-like speakers. (See Figure 7 legend for details.)**

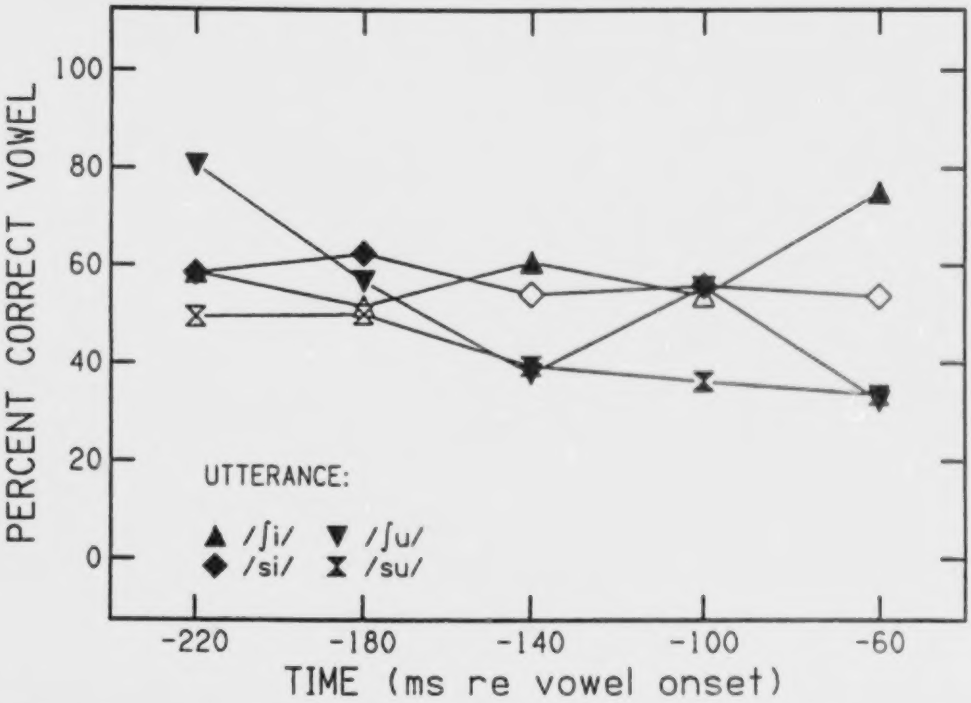


Figure 9. **Mean percent correct vowel identification when the fricative was misidentified for samples from child-like speakers. (See Figure 7 legend for details.)**

EXPERIMENT 3

Having established in Experiment 2 that there was an interaction between fricative and vowel judgments, Experiment 3 was designed to investigate the magnitude of this interaction. To do this, we identified the fricative to listeners. For those vowel judgments based on the F2 resonance, this should have little effect because these judgments were largely independent of fricative identification to begin with. However, the interactions found in Experiment 2 lead us to predict that those syllables which demonstrated a relation between fricative and vowel judgments should show substantial change in accuracy when fricative identity is known.

A. Method

Two child speakers of each age group (one male and one female) and four adult speakers (two male and two female) were randomly selected. Children were not distinguished based on centroid differences for /ʃ/ and /s/ noises because listeners would be told the identity of the fricative. Slices 90 ms in length were extracted from each of the 32 tokens of each speaker, starting 120 ms before vowel onset. Therefore, slices ended 30 ms before vowel onset, the point at which vowel-related changes in fricative F2 are first observed in adult fricative spectra. Separate audiotapes were made for /ʃ/ and for /s/ slices. Stimuli were blocked by speaker, with order of blocks randomized and four practice items starting each block (two slices from the /i/ context, and two from the /u/ context), resulting in tapes consisting of 240 items each (12 speakers × 20 items). Although the two categories of fricatives were presented separately and stimuli were blocked by speaker, listeners were not told whether speakers were children or adults. The order of presentation of the blocks (and therefore of speakers), was randomized, so that adult and child utterances occurred throughout the experiment.

Eight listeners participated in this experiment, half hearing the /ʃ/ tape first, and half hearing the /s/ tape first. The task was simply to write what they thought the vowel context was.

B. Results and Discussion

Figure 10 shows the percent correct vowel identifications for each of the four syllables, divided by speaker group. Superimposed on each column are the most comparable values from Experiment 2, which are the correct vowel identification scores for the -100 ms stimuli.

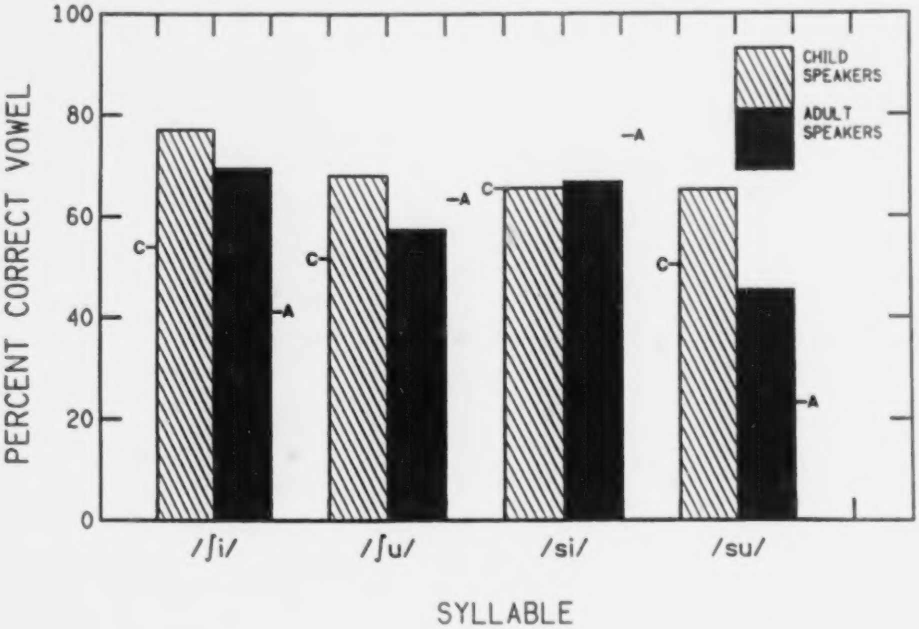


Figure 10. **Mean percent correct vowel identification from Experiment 3, in which the identity of the fricative was known. Tic marks next to each column represent the corresponding value from Experiment 2, in which identity of the fricative was not known.**

For the child speakers, these scores were averaged across the two types of child speakers of Experiment 2 (since we did not distinguish between them in this experiment). Collapsing across syllable type, vowel identification scores in this third experiment were once again more accurate for child (69%) than for adult (60%) speakers ($F(1,7) = 38.30$, $p < .01$). The size of the difference is almost exactly that found in the first two experiments, about 10%, adding further support to the interpretation of this age effect as due to enhanced F2 information for children's samples.

For those syllables which showed an interaction between fricative and vowel judgments in Experiment 2 (that is, /ʃi/ and /su/, which are on the extreme left and right, respectively, in Figure 10), knowing the fricative identity resulted in an average gain in vowel identification scores of 22 percentage points for adults' samples and 17 percentage points for children's samples. For the syllables that did not demonstrate an interaction between fricative and vowel judgments (that is, /ʃu/ and /si/), knowing the identity of the fricative contributed little: Scores for adults' samples actually decreased by 7.5 percentage points, while scores for children's samples improved by 5 percentage points. Thus, in just those cases where removing the source of interaction should have improved performance, performance was improved. Moreover, vowel judgments for all children's samples improved to some extent when fricative identity was given. Apparently in the earlier two experiments, the decreased ability of listeners to accurately judge fricative identity for children's samples actually degraded vowel identification scores somewhat.

GENERAL DISCUSSION

On the whole, we have found the following answers to the questions posed in the Introduction:

1) The greater coarticulation between vowel and fricative gestures reported by Nittrouer et al. (1989) for children's speech is indeed perceivable by adult listeners as phonetic information, and not just as noise.
2) The enhanced coarticulation of children's FV syllables is greater in temporal extent, as well as in magnitude.
3) Perception of a coarticulated phone can include both processes that are dependent on recognition of the prominent phone, as well as processes that are independent of recognition of that phone.

These conclusions serve to support and extend the findings of Nittrouer et al. (1989), and to provide new insights into the processes of speech perception. With children's speech, listeners were able to separate the enhanced vowel information from the degraded consonant information. It cannot be the case that listeners heard an /ʃ/ or an /s/, and then determined vowel context from the allophonic variation of this phonetic segment because, for children's tokens, listeners were often unable to determine whether the fricative was an /ʃ/ or an /s/.

At the same time though, evidence was also found for the idea that phonetic judgments of adjacent segments interact under certain conditions: specifically, in those cases where both segmental judgments are based on the same acoustic information. In fact, other work by one of us (Whalen, 1989) reached the same conclusion. In that study, listeners were asked to identify both the fricative and the vowel of synthetic fricative-vowel syllables. The frequencies of the fricative spectral prominences were deliberately chosen to be ambiguous between /ʃ/ and /s/, and the vocalic formants were chosen to be ambiguous between /i/ and /u/. The fricative noises had no F2, and the vocalic portions had no formant transitions. Consequently, it was impossible to use the fricative F2 to identify the upcoming vowel, or to use the vocalic F2 to identify the preceding fricative. When a fricative noise was identified as /ʃ/, vowel judgments were more often /i/ than when the same noise was identified as /s/. Listeners must have used the fricative spectral prominence to make judgments about both fricative and vowel identity. Thus, the use of the same acoustic information to make phonetic judgments about adjacent segments apparently dictates that these judgments be dependent (i.e., that they interact), while the use of different acoustic information allows those judgments to be independent.

The present results are at variance with those of Sereno et al. (1987), who found no evidence of greater coarticulation in the productions of children. Several reasons for the difference on the acoustic side have been given in Nittrouer et al. (1989), as well as in the Introduction here. On the perceptual side, though, we now have an alternative explanation for at least some of their results. In Sereno et al., the listeners were required to give only a vowel judgment. Thus, for the fricative noises, we do not know from their data whether the same type of interaction that was found in the present study also occurred, namely, that their listeners in fact misheard the fricative in some cases and therefore misheard the vowel as well. Whether or not this happened, there is a further difference between the two studies, namely that Sereno et al. grouped the stimuli of adult and child speakers separately, while ours were always randomized together. Furthermore, the adult stimuli were always presented first in Sereno et al., perhaps leading listeners to adopt a successful strategy for adult sounds that subsequently failed with children's productions. In any case, the issue is of enough importance that we can expect other studies to further illuminate these differences.

One additional result of interest was observed in the present data: The F2 frequency *per se* did not determine perceptual judgments. Nittrouer et al. (1989) showed that the absolute values of children's fricative F2 frequencies were always higher than those of adults. If the identification of vowel context was dependent on this acoustic parameter, all slices from children's samples would have been identified as coming from the /i/ context. The fact that this result was not observed indicates that listeners do not perform direct acoustic-to-phonetic mappings. Neither could listeners have been doing some sort of acoustic normalization of the sort proposed by various authors (e.g., Assman et al., 1982; Gerstman, 1968; Lieberman et al., 1972; Nearey, 1978; Strange et al., 1976; Syrdal & Gopal, 1986). The parameters considered necessary for vocal tract normalization by these proposed algorithms were not usually available in these experiments: Only one of the first two formants was present; F2 frequencies for only two vowels per speaker were present; stimuli were not presented in a carrier phrase; many stimuli were extremely brief in duration and often did not include dynamic information; and usually the stimuli were not blocked by speaker. Instead, the acoustics must provide information about something else. Direct accounts of speech perception (Fowler, 1984; 1986) hold that the acoustic signal directly specifies articulatory events, and all results from the present set of experiments are in keeping with this idea: When gestures were poorly produced, as Nittrouer et al. (1989) found for constriction shape of children's fricatives, perceptual results were poor, when gestures were produced well, as was found to be the case for the enhanced vowel tongue gestures within the children's fricatives, perceptual results were good. Thus, within these extremely brief acoustic slices, some information was apparently available about the size of the vocal tract generally, and about the vowel-related posture of the tongue specifically.

Finally, it seems that the articulatory events perceived are not simply articulatory movements, such as liprounding, but are more global characteristics of the vocal tract, such as volume of the cavity anterior to a fricative constriction. Moreover, when different articulatory gestures affect the same global vocal tract parameter, listeners apparently cannot separate these gestures; however, when different parameters are affected, listeners can. In the present study, both anticipatory liprounding and characteristics of the fricative constriction affected the volume of the cavity anterior to the fricative constriction, and so the frequency of the fricative spectral prominence. Listeners seemed unable to judge how cavity volume, as indicated by the frequency of this prominence, was being affected separately by constriction and by liprounding characteristics. Consequently, if an erroneous decision was made concerning one, an erroneous conclusion was often reached concerning the other. On the other hand, when different articulatory events primarily affected different aspects of the acoustic signal, then listeners apparently could separate these events. So, in the present study, anticipatory lingual gestures affected fricative F2, which is not as important a cue, if one at all, to fricative identity. Listeners could, to a large extent, separate this information concerning vowel identity from information associated with fricative identity. It seems that when simultaneous articulatory events shape different portions of the acoustic signal, listeners divide the signal into temporally overlapping but qualitatively separate units, about which they make independent phonetic judgments. However, when simultaneous articulatory events shape the same part of the acoustic signal, listeners do not divide the signal into qualitatively separate units, but instead make dependent phonetic judgments of each.

In conclusion, we have demonstrated that the enhanced anticipatory vowel coarticulation previously found in children's fricative-vowel syllables is perceptible. Furthermore, children's coarticulation is not only greater in magnitude, but in extent as well. Listeners are able to separate information about this enhanced vowel coarticulation from that of fricative production for children's speech because, to some extent, the two articulatory events affect different acoustic sources of information. However, it was shown

also that judgments of the two phones interact when they are based on the same acoustic information.

REFERENCES

Assman, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71,* 975-989.

Flege, J. E. (1988). Anticipatory and carry-over nasal coarticulation in the speech of children and adults. *Journal of Speech and Hearing Research, 31,* 525-536.

Forrest, K., Weismer, G., Milenkovic, P. M., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84,* 115-123.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36,*359-368.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics 14, 3-28.

Gerstman, L. J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics, AV-16,* 78-80.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives spoken in syllables. *Language and Speech, 1,* 1-7.

LaRiviere, C. J., Winitz, H., and Herriman, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research, 18,* 613-622.

Liberman, A. M., Cooper, F. S., Shankweiler, D. S., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-36.

Lieberman, P., Crelin, E., & Klatt, D. (1972). Phonetic ability and the related anatomy of the newborn, adult human, Neanderthal man and the chimpanzee. *American Anthropology, 74,* 287-307.

Mann, V. A. , & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics, 28,* 213-228.

McGowan, R. S., & Nittrouer, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ / and /s/. *Journal of the Acoustical Society of America, 83,* 229-236.

Nearey, T. (1978). *Phonetic feature systems for vowels.* Doctoral dissertation, Indiana University Linguistics Club, Bloomington.

Nittrouer, S. (1985). *The role of coarticulation in the perception and production of speech by young children (3 to 7 years).* Unpublished doctoral dissertation, City University of New York.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research, 30,* 319-329.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120-132.

Sereno, J. A., Baum, S. R., Marean, G. C., & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *Journal of the Acoustical Society of America,81,* 512-519.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America,70,* 976-984.

Strange, W., Verbrugge, R., Shankweiler, D., and Edman, T. (1976). Consonant context specifies vowel identity. *Journal of the Acoustical Society of America, 60,* 213-224.

Studdert-Kennedy, M. (1976). Speech perception. In N .J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-293). New York:Academic Press.

Studdert-Kennedy, M. (1985). Perceiving phonetic events. In R. Warren & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the First International Conference on Event Perception* (pp. 139-156). Hillsdale, NJ: Lawrence Erlbaum Associates.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America,79,* 1086-1100.

Whalen, D. H. (1981). Effects of vocalic format transitions and vowel quality on the English [s]-[s] boundary. *Journal of the Acoustical Society of America, 69,* 275-282.

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech, 26,* 91-100.

Whalen, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics, 46,* 284-292.

Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America, 70,* 966-975.

FOOTNOTES

[*] *Journal of the Acoustical Society of America,* 1989.

[1] Mean syllable and fricative-noise durations were not significantly different across age groups (mean syllable duration = 298 ms; mean fricative-noise duration = 166 ms). However, children demonstrated more variability in these durations, both across children and among tokens from individual children.

[2] This planned comparison was done because the demonstration of a significant age effect would show only that there was a difference among scores for some of the speaker groups; it would not isolate this difference to one between children and adults. Conversely, the lack of a significant age effect would not necessarily ensure that there was no difference between the four children's groups and the one adult group.

[3] Ideally, one child from each of the four age groups demonstrating greater than a 1-kHz difference in mean centroids between /ʃ/ and /s/ would have been selected, but software constraints limited us to ten speakers in Experiment 2. There was an absolute upper limit on the number of slices which could be included, and the large number of slices extracted from each individual token quickly brought us near this limit. We wanted to include four adult speakers to avoid the great inequality between the numbers of adult and child speakers which existed in the first experiment, so we were limited to six child speakers. We decided that it would be most reasonable to select the larger number of child speakers who demonstrated a small difference in centroids between /ʃ/ and /s/ because all adult speakers demonstrated greater than a 1-kHz difference, and the possibility existed that children demonstrating a large centroid difference would perform identically to adults.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 194-219

"Targetless" Schwa: An Articulatory Analysis*

Catherine P. Browman and Louis Goldstein†

1. INTRODUCTION

One of the major goals for a theory of phonetic and phonological structure is to be able to account for the (apparent) contextual variation of phonological units in as general and simple a way as possible. While it is always possible to state some pattern of variation using a special "low-level" rule that changes the specification of some unit, recent approaches have attempted to avoid stipulating such rules, and instead propose that variation is often the *consequence* of how the phonological units, properly defined, are organized. Two types of organization have been suggested that lead to the natural emergence of certain types of variation. One is that invariantly specified phonetic units may overlap in time, i.e., they may be co-produced (e.g., Bell-Berti & Harris, 1981; Browman & Goldstein, 1987; Fowler, 1977; 1981a,b; Liberman & Mattingly, 1985), so that the overall tract shape and acoustic consequences of these co-produced units will reflect their combined influence. A second is that a given phonetic unit may be unspecified for some dimension(s) (e.g., Keating, 1988; Öhman, 1966), so that the apparent variation along that dimension is due to continuous trajectories between neighboring units' specifications for that dimension.

A particularly interesting case of contextual variation involves reduced (schwa) vowels in English. Investigations have shown that these vowels are particularly malleable: they take on the acoustic (Fowler, 1981a) and articulatory (e.g., Alfonso & Baer, 1982) properties of neighboring vowels. While Fowler (1981a) has analyzed this variation as emerging from the co-production of the reduced vowels and a neighboring stressed vowel, it might also be the case that schwa is completely unspecified for tongue position. This would be consistent with analyses of formant trajectories for medial schwa in trisyllabic sequences (Magen, 1989) that have shown that F2 moves (roughly continuously) from a value dominated by the preceding vowel (at onset) to one dominated by the following vowel (at offset). Such an analysis would also be consistent with the phonological analysis of schwa in French (Anderson, 1982) as an empty nucleus slot. It is possible (although this is not Anderson's analysis), that the empty nucleus is never filled in by any specification, but rather there is a specified "interval" of time between two full vowels in which the tongue continuously moves from one vowel to another.

The computational gestural model being developed at Haskins Laboratories (e.g., Browman & Goldstein, 1987; Browman, Goldstein, Saltzman, & Smith, 1986; Saltzman, Goldstein, Browman, & Rubin, 1988) can serve as a useful vehicle for testing these (and other) hypotheses about the phonetic/phonological structure of utterances with such reduced schwa vowels. As we will see, it is possible to provide a simple, abstract representation of such utterances in terms of gestures and their organization that can yield the variable patterns of articulatory behavior and acoustic consequences that are observed in these utterances.

The basic phonetic/phonological unit within our model is the *gesture*, which involves the formation (and release) of a linguistically significant constriction within a particular vocal tract subsystem. Each gesture is modeled as a dynamical system (or set of systems) that regulates the time-varying coordination of individual articulators in performing these constriction tasks (Saltzman, 1986). The dimensions along which the vocal tract goals for constrictions can be specified are called *tract variables*, and are shown in the left-hand column of Figure 1. Oral constriction gestures are defined in terms of pairs of these tract variables, one for constriction location, one for constriction degree.

Our thanks to Carol Fowler and Doug Whalen for critiquing the preliminary version of this paper. This work was supported by NSF grant BNS 8820099 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

	tract variable	articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum

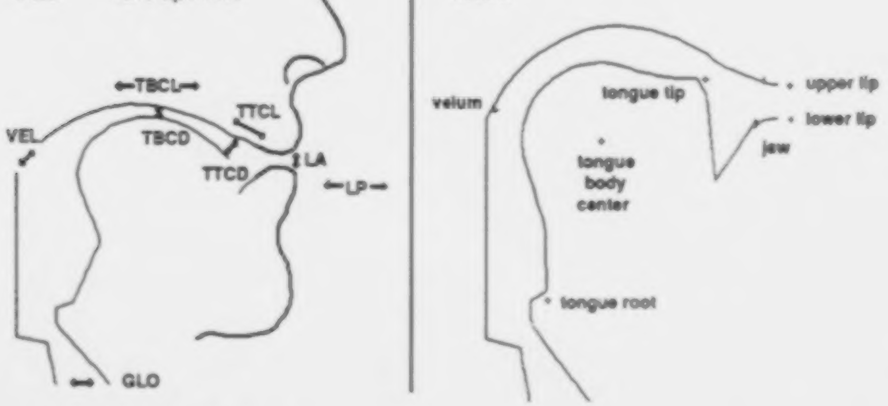


Figure 1. Tract variables and associated articulators.

The right-hand side of the figure shows the individual articulatory degrees of freedom whose motions contribute to the corresponding tract variable.

The computational system we have developed, sketched in Figure 2, (Browman & Goldstein, 1987; Saltzman, Goldstein, Browman, & Rubin, 1988) provides a representation for arbitrary (English) input utterances in terms of such gestural units and their organization over time, called the *gestural score*. The layout of the gestural score is based on the principles of intergestural phasing (Browman & Goldstein, 1987) specified in the *linguistic gestural model*. The gestural score is input to the *task dynamic model* (Saltzman, 1986; Saltzman & Kelso, 1987), which calculates the patterns of articulator motion that result from the set of active gestural units. The articulatory movements produced by the task dynamic model are then input to an articulatory synthesizer (Rubin, Baer, & Mermelstein, 1981) to calculate an output speech waveform. The operation of the task dynamic model is assumed to be "universal." (In fact, it is not even specific to speech, having originally been developed, Saltzman & Kelso, 1987, to describe coordinated reaching movements). Thus, all of the language-particular phonetic/phonological structure must reside in the gestural score—in the dynamic parameter values of individual gestures, or in their relative timing. Given this constraint, it is possible to test the adequacy of some particular hypothesis about phonetic structure, as embodied in a particular gestural score, by using the model to generate the articulatory motions and comparing these to observed articulatory data.

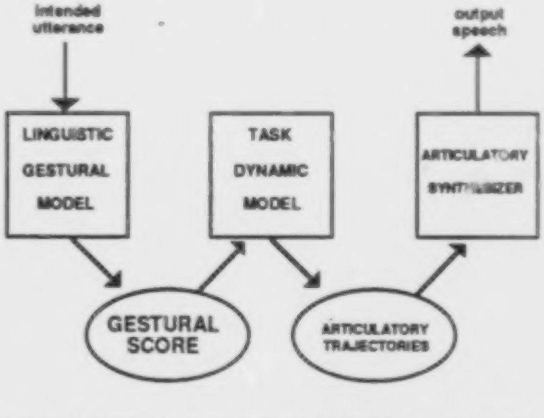


Figure 2. Overview of GEST: gestural computational model.

The computational model can thus be seen as a tool for evaluating the articulatory (and acoustic) consequences of hypothesized aspects of gestural structure. In particular, it is well suited for evaluating the consequences of the organizational properties discussed above: (1) underspecification and (2) temporal overlap. Gestural structures are inherently underspecified in the sense that there are intervals of time during which the value of a given tract variable is not being controlled by the system: only when a gesture defined along that tract variable is active is such control in place. This underspecification can be see in Figure 3, which shows the gestural score for the utterance /paam/ (ARPAbet is used throughout the text, for phonetic transcription).[1] Here, the shaded boxes indicate the gestures, and are superimposed on the tract variable time functions produced when the gestural score is input to the task dynamic model. The horizontal dimension of the shaded boxes indicates the intervals of time during which each of the gestural units is active, while the height of the boxes corresponds to the "target" or equilibrium position parameter of the second-order dynamical regime.

Note that during the activation interval of the initial bilabial closure gesture, Lip Aperture (vertical distance between the two lips) gradually decreases, until it approaches the regime's target. However, even after the regime is turned off, LA shows changes over time. Such "passive" tract variable changes result from two sources: (1) the participation of one of the (uncontrolled) tract variable's articulators in some other tract variable which *is* under active gestural control, and (2) an articulator-specific "neutral" or "rest" regime, that takes control of any articulator which is not currently active in any gesture. For example, in the LA case shown here, the jaw contributes to the TBCD gesture (for the vowel) by lowering, and this has the side effect of increasing LA. In addition, the upper and lower lips are not involved in any active gesture, and so move toward their neutral positions with respect to the upper and lower teeth, thus further contributing to an increase in LA. Thus, the geometric structure of the model itself, (together with the set of articulator neutral values) predicts a specific, well-behaved time function for a given tract variable, even when it is not being controlled. Uncontrolled behavior need not be stipulated in any way. This feature of the model is important to being able to test the hypothesis that schwa may not involve an active gesture at all.

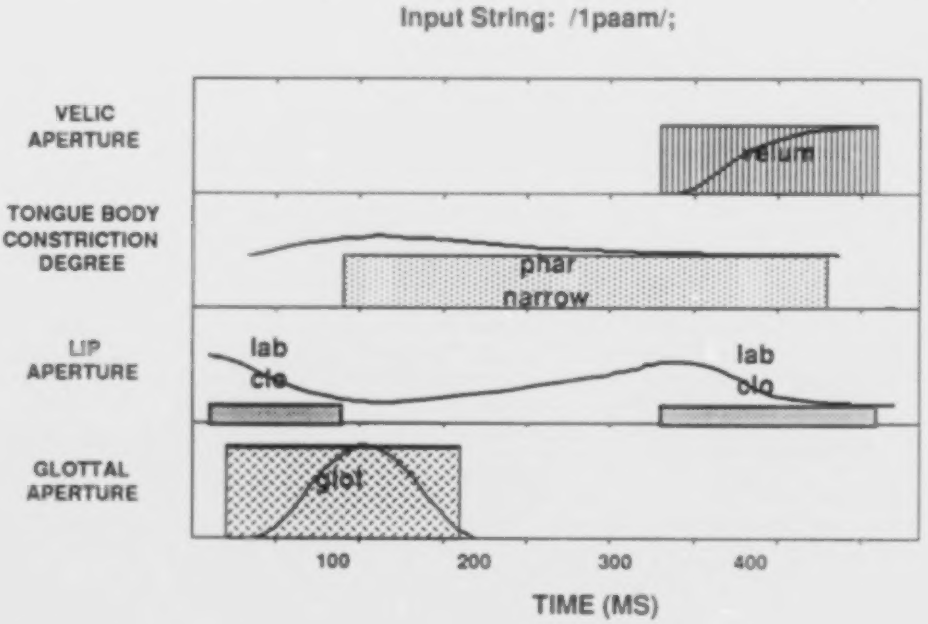


Figure 3. Gestural score and generated motion variables for /paam/.

The second useful aspect of the model is the ability to predict consequences of the temporal overlap of gestures, i.e., intervals during which there is more than one concurrently active gesture. Browman and Goldstein (1987) have shown that the model predicts different consequences of temporal overlap, depending on whether the overlapping gestures involve the same or different tract variables, and that these different consequences can actually be observed in allophonic variations and "casual speech" alternations. Of particular importance to analyzing schwa is the shared tract variable case, since we will be interested in the effects of overlap between an active schwa gesture (if any) and the preceding or following vowel gesture, all of which would involve the Tongue Body tract variables (TBCD, TBCL). In this case, the dynamic parameter values for the overlapping gestures are "blended," according to a competitive blending dynamics (Saltzman, Goldstein, Browman, & Rubin, 1988; Saltzman & Munhall, 1989). In the examples we will be examining, the blending will have the effect of averaging the parameter values. Thus, if both gestures were coextensive for their entire activation intervals, neither target value would be achieved, rather, the value of the tract variable at the end would be the average of their targets.

In this paper, our strategy is to analyze movements of the tongue in utterances with schwa to determine if the patterns observed provide evidence for a specific schwa tongue target. Based on this analysis, specific hypotheses about the gestural overlap in utterances with schwa are then tested by means of computer simulations using the gestural model described above.

2. ANALYSIS OF ARTICULATORY DATA

Using data from the Tokyo X-ray archive (Miller & Fujimura, 1982), we analyzed /pV1pax'pV2pax/ utterances produced by a speaker of American English, where V1 and V2 were all possible combinations of /iy,eh,aa,ah,uw/. Utterances were read in short lists of seven or eight items, each of which had the same V1 and different V2s. One token (never the initial or final item in a list) of each of the twenty-five utterance types was analyzed. The microbeam data tracks the motion of five pellets in the mid-sagittal plane. Pellets were located on the lower lip (L), the lower incisor for jaw movement (J), and the midline of the tongue: one approximately at the tongue blade (B), one at the middle of the tongue dorsum (M), and one at the rear of the tongue dorsum (R).

Ideally, we would use the information in tongue pellet trajectories to infer a time-varying representation of the tongue in terms of the dimensions in which vowel gesture targets are defined, e.g., for our model (or for Wood, 1982), location and degree of tongue body constriction. (For Ladefoged & Lindau, 1989, the specifications would be rather in terms of formant frequencies linked to the factors of front-raising and back-raising for the tongue). Since this kind of transformation cannot currently be performed with confidence, we decided to describe the vowels directly in terms of the tongue pellet positions. As the tongue blade pellet (B) was observed to be largely redundant in these utterances (not surprising, since they involve only vowels and bilabial consonants), we chose to measure the horizontal (X) and vertical (Y) positions (in Euclidean space) for the M and R pellets for each vowel. While not ideal in one sense, the procedure has the advantage that the analysis does not make any a priori assumptions about the correct parameterization of tongue shape.

The first step was to find appropriate time points at which to measure the position of the pellets for each vowel. The time course of each tongue pellet dimension (MX, MY, RX, RY) was analyzed by means of an algorithm that detected displacement extrema (peaks and valleys). To the extent that there is a characteristic pellet value associated with a given vowel, we may expect to see such a displacement extremum, that is, movement towards some value, then away again. The algorithm employed a noise level of one x-ray grid unit (= approximately .33 mm); thus, movements of a single unit in one direction and back again did not constitute extrema. Only the interval that included the full vowels and the medial schwa was analyzed; final schwas were not analyzed. In general, an extremum was found that coincided with each full vowel, for each pellet dimension, while such an extremum was missing for schwa in over half the cases. The pellet positions at these extrema were used as the basic measurements for each vowel. In cases where a particular pellet dimension had no extremum associated with a vowel, a reference point was chosen that corresponded to the time of an extremum of one of the other pellets. In general, MY was the source of these reference points for full vowels, and RY was the source for schwa, as these were dimensions that showed the fewest missing extrema. After the application of this algorithm, each vowel in each utterance was categorized by the value at a single reference point for each of the four pellet dimensions. Since points were chosen by looking only at data from the tongue pellets themselves, these are referred to as the "tongue" reference points.

To illustrate this procedure, Figure 4a shows the time courses of the M, R, and L pellets (only vertical for L) for the utterance /piypax'piypax/ with the extrema marked with dashed lines. The acoustic waveform is displayed at the top. For orientation, note that there are four displacement peaks marked for LY, corresponding to the raising of the lower lip for the four bilabial closure gestures for the consonants. Between these peaks three valleys are marked, corresponding to the opening of the lips for the three vowels. For MX, MY, and RX, an extremum was found associated with each of the full vowels and the medial schwa. For RY, a peak was found for schwa, but not for V1. While there is a valley detected following the peak for schwa, it occurs during the consonant closure interval, and therefore is not treated as associated with V2. Figure 4b shows the same utterance with the complete set of "tongue" reference points used to characterize each vowel. Reference points that have been copied from other pellets (MY in both cases) are shown as solid lines. Note that the consonant closure interval extremum has been deleted.

Figure 5 shows the same displays for the utterance /piypax'paapax/. Note that, in (a), extrema are missing for schwa for MX, MY and RX. This is typical of cases in which there is a large pellet displacement between V1 and V2. The trajectory associated with such a displacement moves from V1 to V2, with no intervening extremum (or even, in some cases, no "flattening" of the curve).

As can be seen Figures 4 and 5, the reference points during the schwa tend to be relatively late in its acoustic duration. As we will be evaluating the relative contributions of V1 and V2 in determining the pellet positions for schwa, we decided to also use a reference point earlier in the schwa. To obtain such a point, we used the valley associated with the lower lip for the schwa—that is, approximately the point at which the lip opening is maximal. This point, called the "lip" reference, typically occurs earlier in the (acoustic) vowel duration than the "tongue" reference point, as can be seen in Figures 4 and 5. Another advantage of the "lip" reference point is that all tongue pellets are measured at the same moment in time. Choosing points at different times for different dimensions might result in an apparent differential influence of V1 and V2 across dimensions. The "lip" reference point was used only for the medial schwa; the full vowel references used for comparison with the "lip" schwa point were the same as those used for comparison with the "tongue" schwa point, and were determined using the tongue extremum algorithm described above.

2.1 Results

Figure 6 shows the positions of the M (on the right) and R (on the left) pellets for the full vowels, plotted in the mid-sagittal plane such that the speaker is assumed to be facing to the right. The ten points for a given vowel are enclosed in an ellipse indicating their principal components (two standard deviations along each axis). The tongue shapes implied by these pellet positions are consistent with cinefluorographic data for English vowels (e.g., Harshman, Ladefoged, & Goldstein, 1977; Neary, 1980; Perkell, 1969). For example, /iy/ is known to involve a shape in which the front of the tongue is bunched forward and up toward the hard palate, compared, for example to /eh/, which has a relatively unconstricted shape. This fronting can be seen in both pellets. In fact, over all vowels, the horizontal components of the motion of the two pellets are highly correlated (r=.939 in the full vowel data, between RX and MX over the 25 utterances). The raising for /iy/ can be seen in M (on the right), but not in R, for which /iy/ is low, lower for example than /aa/. The low position of the back of the tongue dorsum for /iy/ can, in fact, be seen in mid-sagittal cinefluorographic data. Superimposed tongue surfaces for different English vowels (e.g., Ladefoged, 1982) reveals that the curves for /iy/ and /aa/ cross somewhere in the upper pharyngeal region, so that in front of this point, /iy/ is higher than /aa/, while in back of this point, /aa/ is higher. This suggests that the R pellet in the current experiment is far enough back to be behind this cross-over point. /uw/ involves raising of the rear of the tongue dorsum (toward the soft palate), which is here reflected in the raising of both the R and M pellets. In general, the vertical components of the two pellets are uncorrelated across the set of vowels as a whole (r=.020), reflecting, perhaps the operation of two independent factors such as "front-raising" and "back-raising" (Ladefoged, 1980).

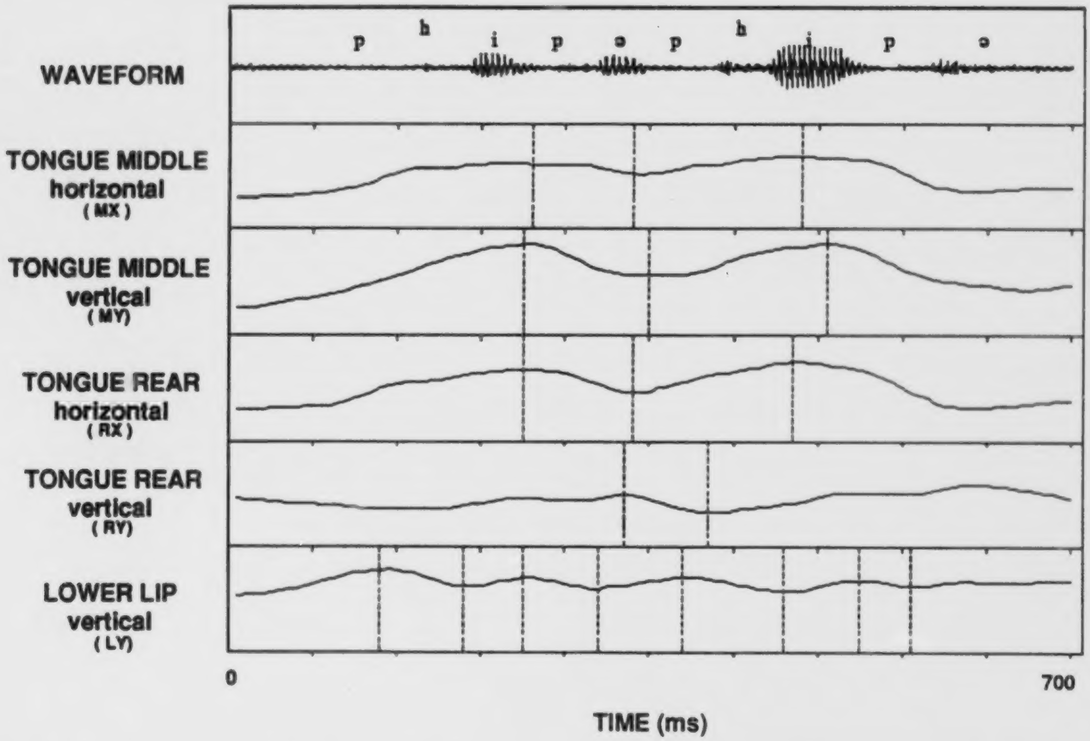
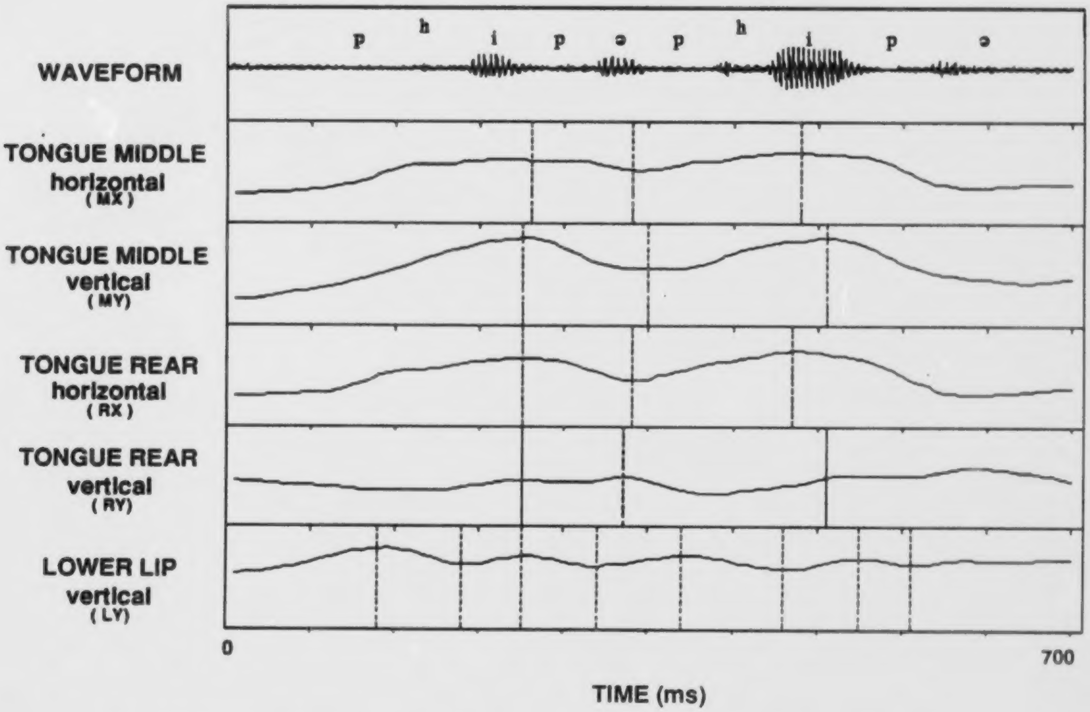


Figure 4. Pellet time traces for /piypax'piypax/ (IPA symbols are used in the figure). *(a)* Position extrema indicated by dashed lines.



(b) "Tongue" reference points indicated by dashed and solid lines (for Middle and Rear pellets).

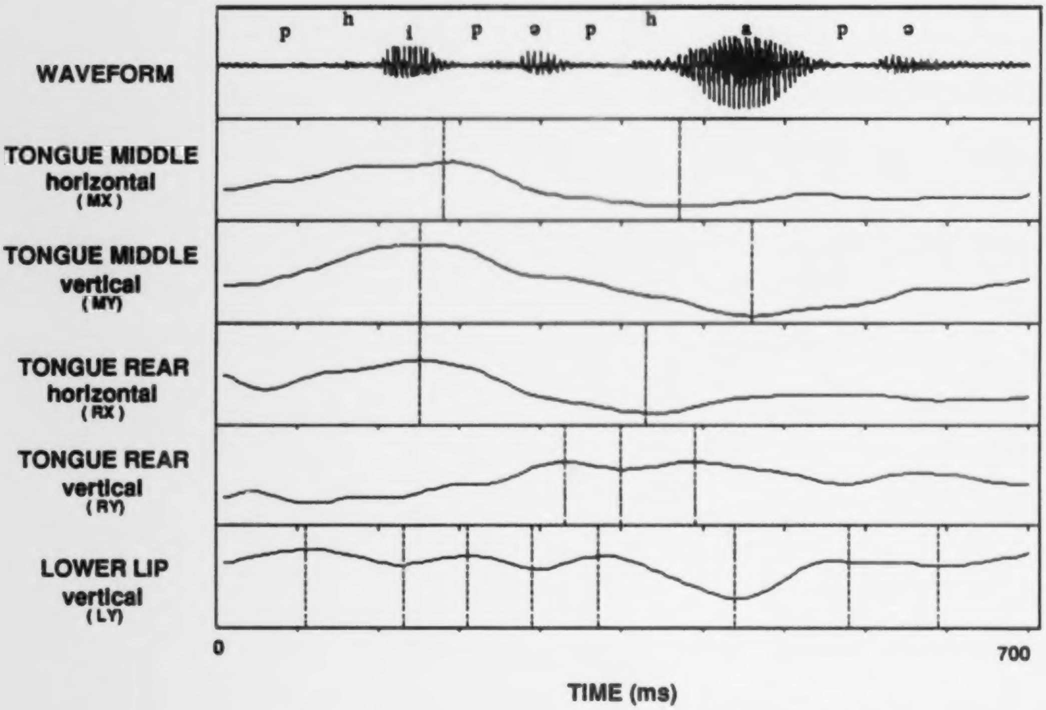
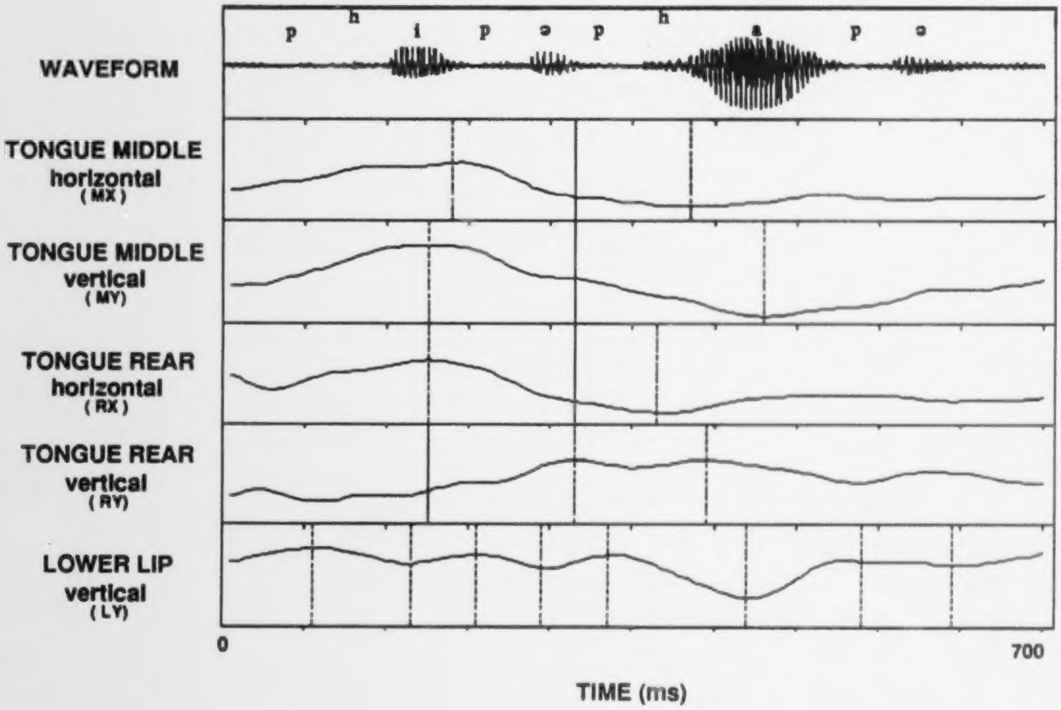


Figure 5. **Pellet time traces for /piypax'paapax/ (IPA symbols are used in the figure).** ***(a)*** **Position extrema indicated by dashed lines.**



(b) **"Tongue" reference points indicated by dashed and solid lines (for Middle and Rear pellets).**

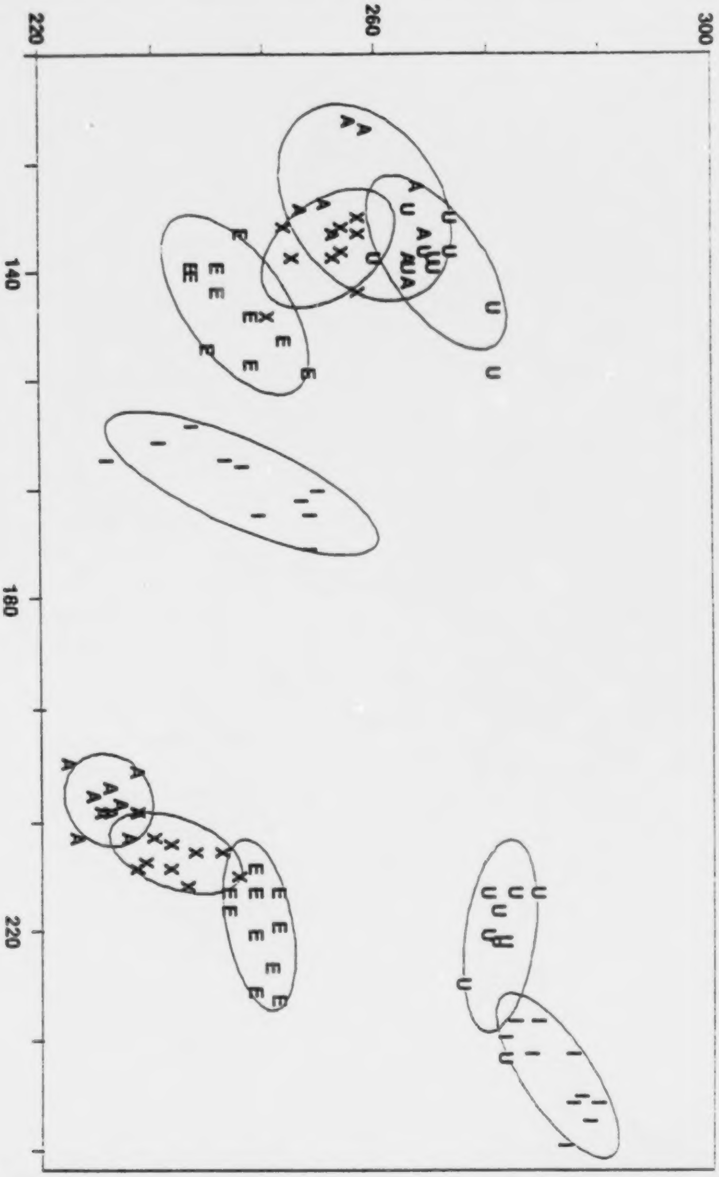


Figure 6. **Pellet positions for full vowels, displayed in mid-sagittal plane with head facing to the right: Mid pellets on the right, Rear pellets on the left. The ellipses indicate two standard deviations along axes determined by principal component analysis. Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are X-ray units (= .33 mm).**

The pellet positions for schwa, using the "tongue" reference points, are shown in the same mid-sagittal plane in Figure 7, with the full vowel ellipses added for reference. The points are labelled by the identity of the following vowel (V2) in (a) and by the preceding vowel (V1) in (b). Figure 8 shows the parallel figure for schwa measurements at the "lip" reference point. In both figures, note that the range of variation for schwa is less than the range of variation across the entire vowel space, but greater than the variation for any single full vowel. Variation in MY is particularly large compared to MY variation for any full vowel. Also, while the distribution of the R pellet positions appears to center around the value for unreduced /ah/, which might be thought to be a target for schwa, this is clearly not the case for the M pellet, where the schwa values seem to center around the region just above /eh/. For both pellets, the schwa values are found in the center of the region occupied by the full vowels. In fact, this relationship turns out to be quite precise.

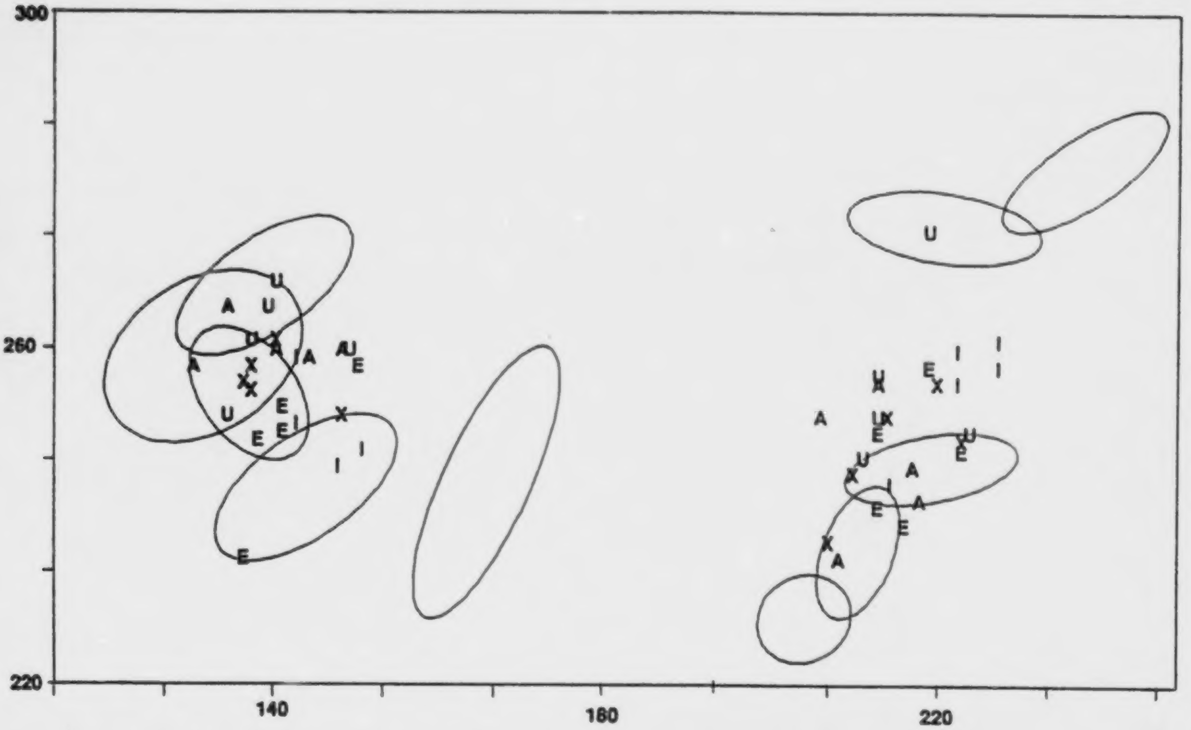
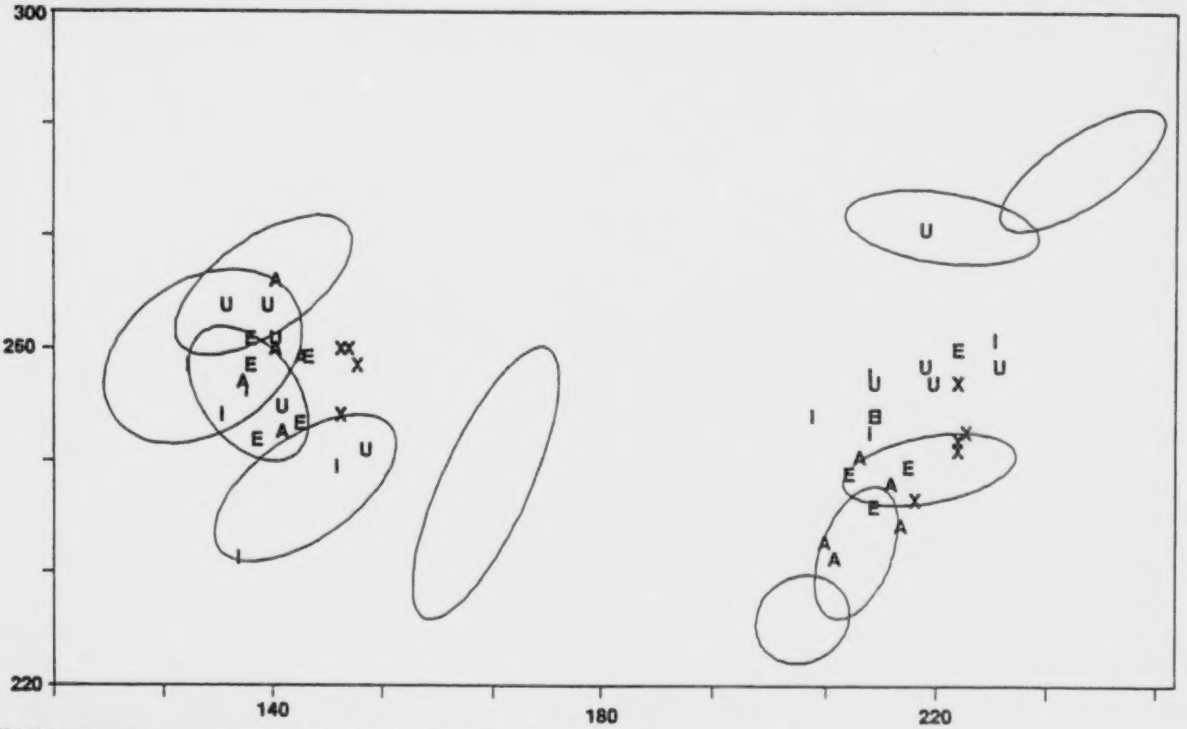


Figure 7. **Pellet positions for schwa at "tongue" reference points, displayed in right-facing mid-sagittal plane as in Figure 6. The ellipses are from the full vowels (Figure 6), for comparison. Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are X-ray units (= .33 mm).** ***(a)*** **Schwa pellet positions labelled by the identity of the following vowel (V2).**



(b) **Schwa pellet positions labelled by the identity of the preceding vowel (V1).**

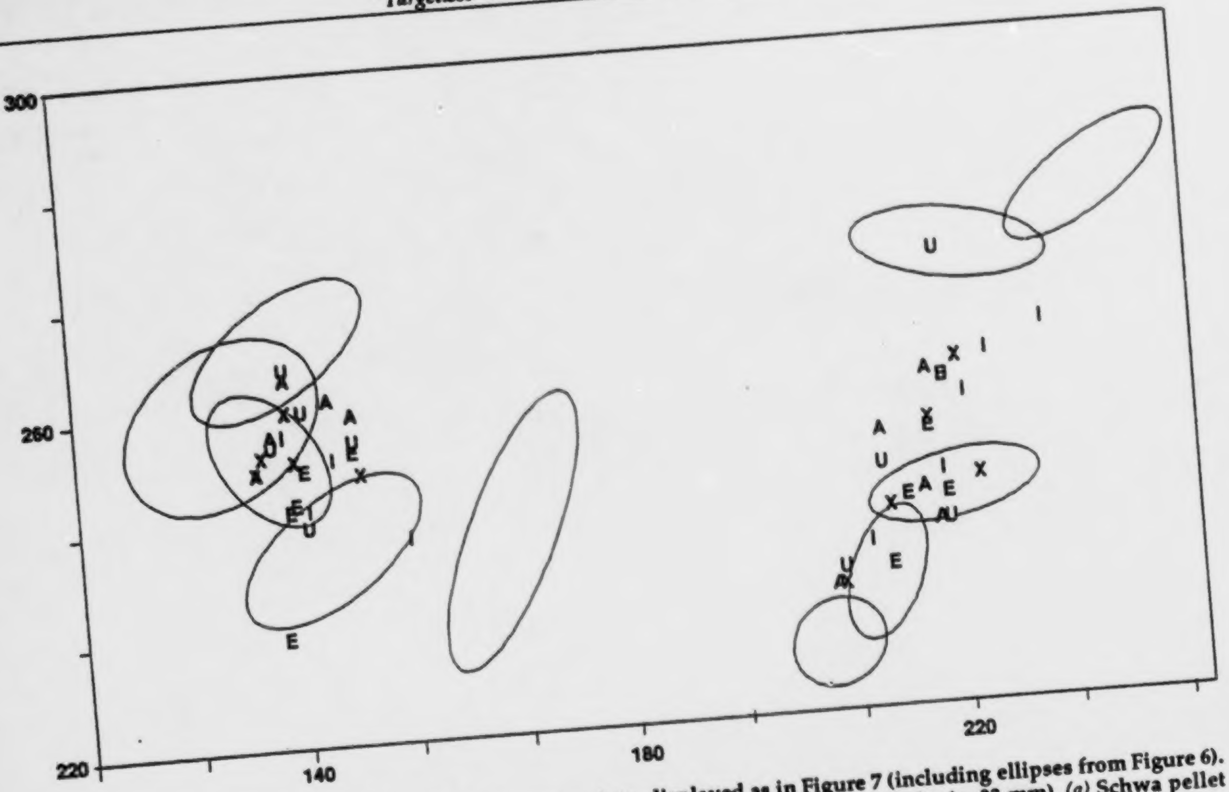
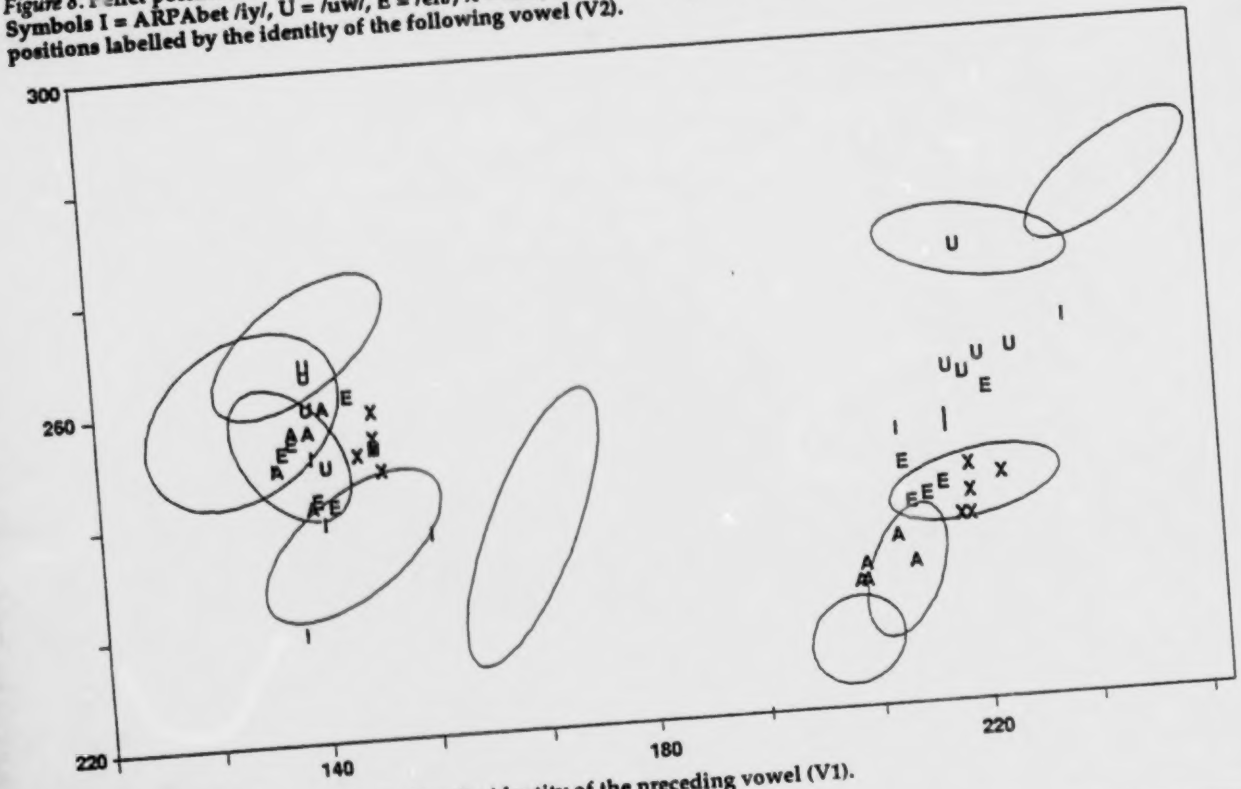


Figure 8. **Pellet positions for schwa at "lip" reference points, displayed as in Figure 7 (including ellipses from Figure 6). Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are X-ray units (= .33 mm).** ***(a)*** **Schwa pellet positions labelled by the identity of the following vowel (V2).**



(b) **Schwa pellet positions labelled by the identity of the preceding vowel (V1).**

Figure 9 shows the mean pellet positions for each full vowel and for schwa ("lip" and "tongue" reference points give the same overall means), as well as the grand mean of pellet positions across all full vowels, marked by a circle. The mean pellet positions for the schwa lie almost exactly on top of the grand mean for both the M and R pellets. This pattern of distribution of schwa points is exactly what would be expected if there were no independent target for schwa but rather a continuous tongue trajectory from V1 to V2. Given all possible combinations of trajectory endpoints (V1 and V2), we would expect the mean value of a point located at (roughly) the midpoint of these 25 trajectories to have the same value as the mean of the endpoints themselves.

If it is indeed the case that the schwa can be described as a targetless point on the continuous trajectory from V1 to V2, then we would expect that the schwa pellet positions could be predicted from knowledge of V1 and V2 positions alone, with no independent contribution of schwa. To test this, we performed stepwise multiple linear regression analyses on all possible subsets of the predictors of V1 position, V2 position, and an independent schwa factor, to determine which (linear) combinations of these three predictors best predicted the position of a given pellet dimension during schwa. The analysis finds the values of the *b* coefficients and the constant *k*, in equations like (1) below, that give the best prediction of the actual schwa values.

(1) schwa(predicted) = b1*V1 + b2*V2 + k

The *stepwise* procedure means that variables are added into an equation such as (1) one at a time, in the order of their importance to the prediction. The procedure was done separately for equations with and without the constant term *k* (using BMDP2R and 9R). For those analyses containing the constant term (which is the y-intercept), k represents an independent schwa contribution to the pellet position—when it is the only predictor term, it is the mean for schwa. Those analyses without the constant term (performed using 9R) enabled the contributions of V1 and V2 to be determined in the absence of this schwa component. Analyses were performed separately for each pellet dimension, and for the "tongue" and "lip" reference points.

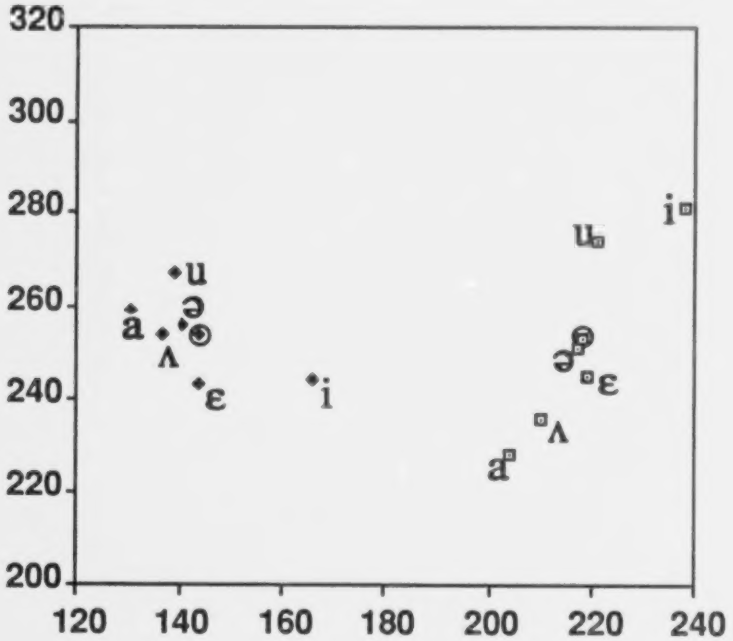


Figure 9. Mean pellet positions for full vowels and schwa, displayed in right-facing mid-sagittal plane as in Figure 6. Phonetic symbols are indicated using IPA; the grand mean of all the full vowels is indicated by a circled square. Units are X-ray units (= .33 mm).

The results for the "tongue" points are shown in the right-hand column of Table 1. For each pellet, the various combinations of terms included in the equation are rank ordered according to the standard error of the schwa prediction for that combination, the smallest error shown at the top.

Table 1. *Regression results for X-ray data.*

	"lip" reference point			"tongue" reference point	
	Terms	*Standard Error*		*Terms*	*Standard Error*
MX	k+v1+v2	4.6	MX	k+v2+v1	4.5
	k+v1	5.5		k+v2	4.6
	v1+v2	5.6		k	6.2
	k	6.3		v2+v1	6.4
	v1	8.6		v2	10.8
MY	k+v1+v2	4.2	MY	k+v1+v2	4.7
	k+v1	5.0		k+v1	6.4
	v1+v2	7.6		k	8.7
	k	10.3		v1+v2	9.1
	v1	11.9		v1	15.1
RX	k+v2+v1	3.8	RX	k+v2+v1	4.0
	k+v2	3.9		k+v2	4.0
	k	4.8		k	5.8
	v1+v2	7.0		v2+v1	7.7
	v1	11.7		v2	10.6
RY	k+v2+v1	4.1	RY	k+v2+v1	4.5
	v2+v1	4.9		k+v2	4.6
	k+v2	5.1		v2+v1	5.3
	k	6.8		v2	6.2
	v2	7.0		k	7.1

In all cases, the equation with all three terms gave the least error (which is necessarily true). Interestingly, however, for MX, RX and RY, the prediction using the constant and V2 differed only trivially from that using all three variables. This indicates that V1 does not contribute substantially to schwa pellet positions at the "tongue" reference point. In addition, it indicates that an independent schwa component is important to the prediction, because V2 alone, or in combination with V1, gives worse prediction than V2 plus k. For MY, all three terms seem to be important—removing any one of them increases the error. Moreover, the second-best prediction involves deleting the V2 term, rather than V1. [The reduced efficacy of V2 (and increased efficacy of V1) in predicting the MY value of schwa may be due, in part, to the peak determination algorithm employed. When V1 or V2 was /aa/ or /ah/, the criteria selected a point for MY that tended to be much later in the vowel than the point chosen for the other pellet dimensions (Figure 5b gives an example of this). Thus, for V2, the point chosen for MY is much further in time from the schwa point than is the case for the other dimensions, while for V1, the point chosen is often closer in time to the schwa point.]

The overall pattern of results can be seen graphically in Figure 10. Each panel shows the relation between "tongue" pellet positions for schwa and the full vowels: V1 in the top row and V2 in the bottom row, with a different pellet represented in each column. The points in the top row represent the pellet positions for the utterances with the indicated initial vowel (averaged across five utterances, each with a different final vowel), while the bottom row shows the average for the five utterances with the indicated final vowel. The relation between schwa and V2 is quite systematic—for every pellet, the lines do not cross in any of the panels of the bottom row—while for V1 (in the top row), the relationship is only systematic for RY (and somewhat for MY).

Turning now to the "lip" reference points, regression results for these points are found in the left-hand column of Table 1. The best prediction again involves all three terms, but here, in every case except RX, the best two-term prediction does substantially worse. Thus V1, which was effectively irrelevant at the "tongue" point, does contribute to the schwa position at this earlier "lip" point. In fact, for these three pellets, the second best prediction combination always involves V1 (with either V2 or k as the second term). This pattern of results can be confirmed graphically in Figure 11. Comparing the V1 effects sketched in the top row of panels in Figures 10 and 11, notice that there is more spread between the schwa pellets at the "lip" point (Figure 11) than at the "tongue" point (Figure 10). This indicates that the schwa pellet was more affected by V1 at the "lip" point. There is also somewhat less crossover for MX and RX in the "lip" figure, indicating increased systematicity of the V1 effect.

In summary, it appears that the tongue position associated with medial schwa cannot be treated simply as an intermediate point on a direct tongue trajectory from V1 to V2. Instead, there is evidence that this V1-V2 trajectory is warped by an independent schwa component.

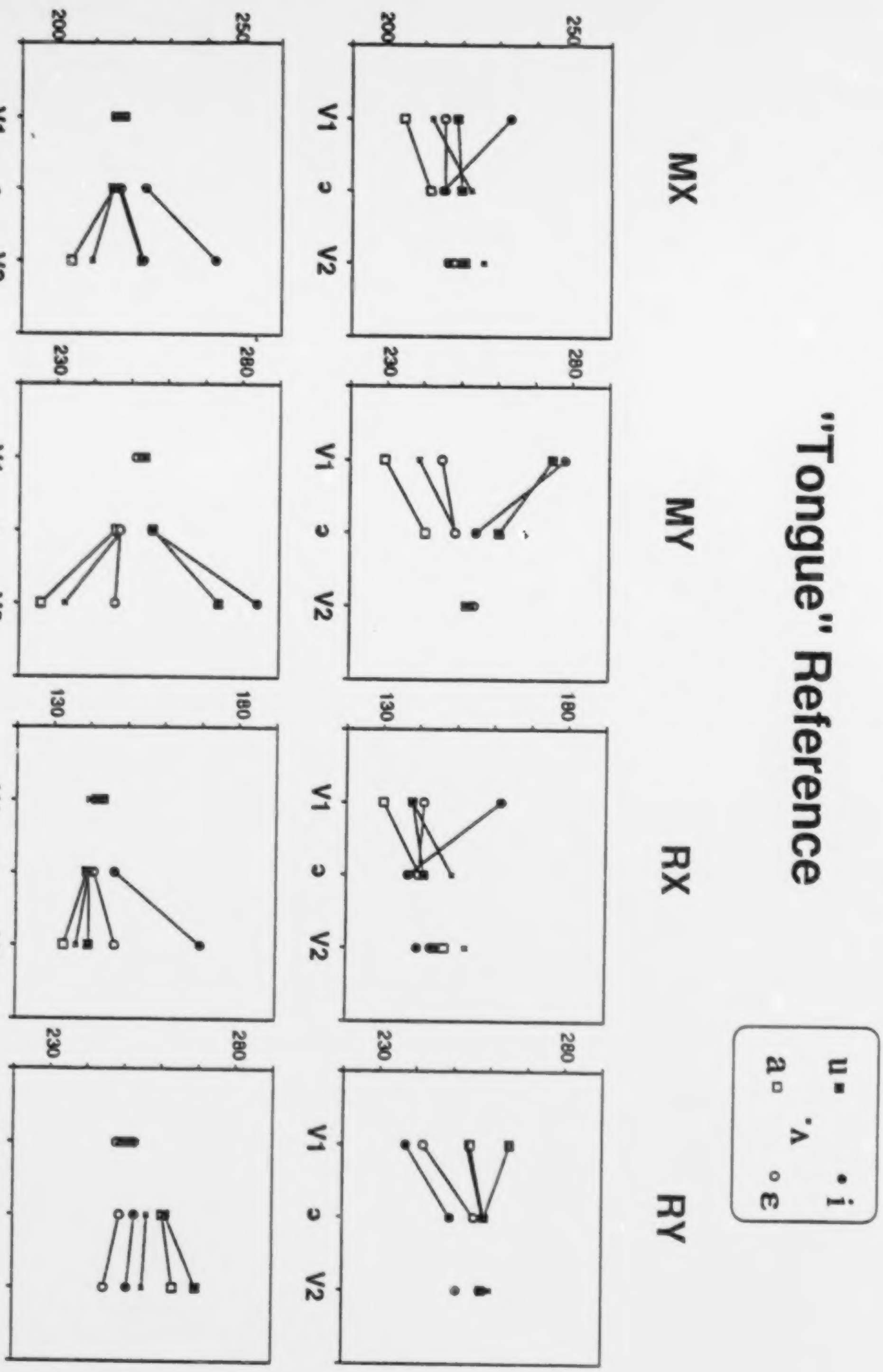


Figure 10. **Relation between full vowel pellet positions and "tongue" pellet positions for schwa. The top row displays the pellet positions for utterances with the indicated initial vowels, averaged across 5 utterances (each with a different final vowel). The bottom row displays the averaged pellet positions for utterances with the indicated final vowels. Units are X-ray units (= .33 mm).**

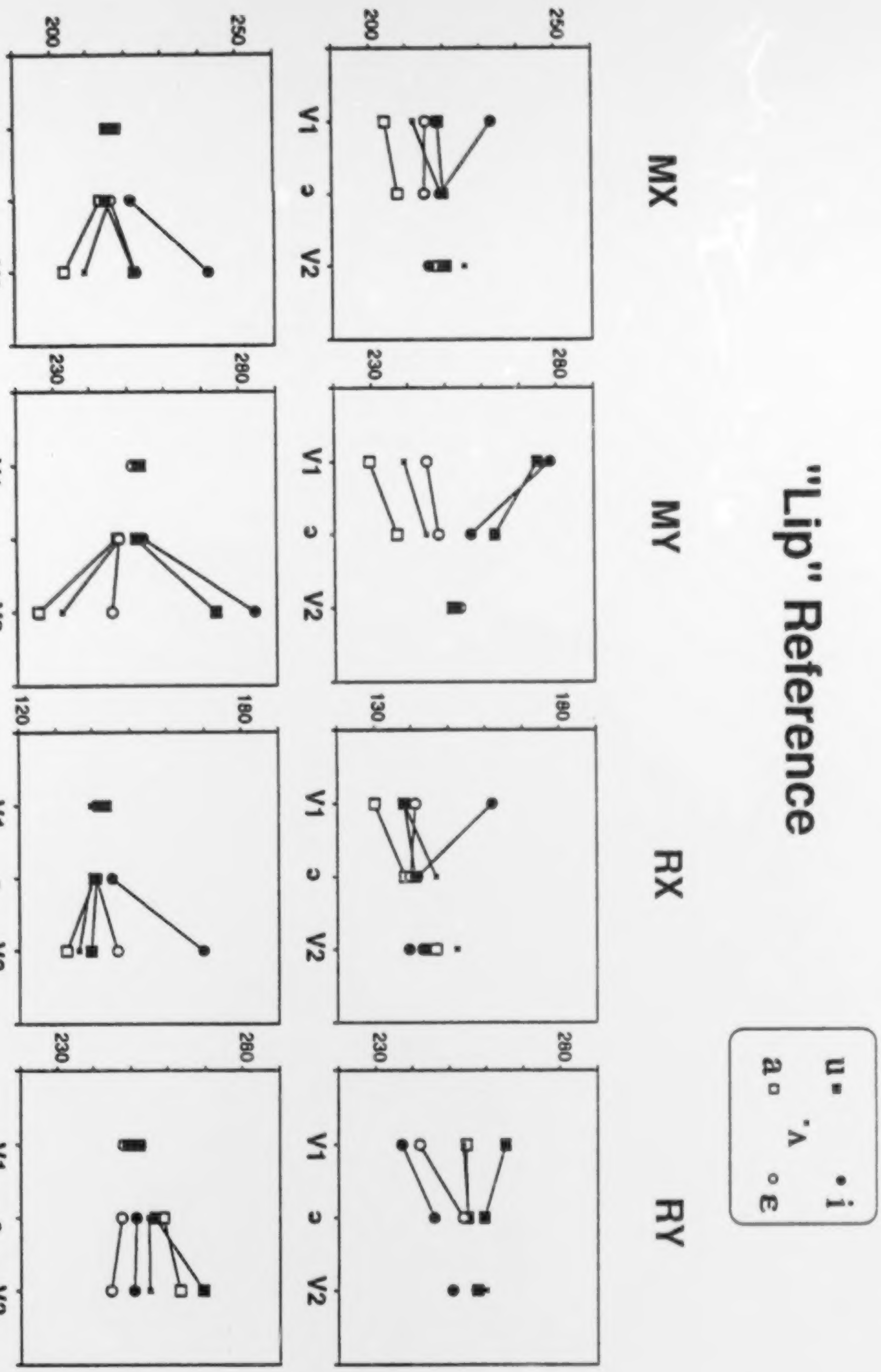


Figure 11. **Relation between full vowel pellet positions and "lip" pellet positions for schwa. The top row displays the pellet positions for utterances with the indicated initial vowels, averaged across 5 utterances (each with a different final vowel). The bottom row displays the averaged pellet positions for utterances with the indicated final vowels. Units are X-ray units (= .33 mm).**

The importance of this warping can be seen, in particular, in utterances where V1 and V2 are identical (or have identical values on a particular pellet dimension). For example, returning to the utterance /piypax'piypax/ in Figure 4, we can clearly see (in MX, MY and RX) that there is definitely movement of the tongue away from the position for /iy/ between the V1 and V2. This effect is most pronounced for /iy/. For example, for MY, the prediction error for the equation without a constant is worse for /piypax'piypax/ than for any other utterance (followed closely by utterances combining /iy/ and /uw/; MY is very similar for /iy/ and /uw/). Yet, it may be inappropriate to consider this warping to be the result of a target specific to schwa, since, as we saw earlier, the mean tongue position for schwa is indistinguishable from the mean position of the tongue across all vowels. Rather the schwa seems to involve a warping of the trajectory toward an overall average or neutral tongue position. Finally, we saw that V1 and V2 affect schwa position differentially at two points in time. The influence of the V1 endpoint can only be seen at the "lip" point, relatively early in the schwa, while V2 influence can be seen throughout. In the next section, we propose a particular model gestural structure for these utterances, and show that it can account for the various patterns that we have observed.

3. ANALYSIS OF SIMULATIONS

Within the linguistic gestural model of Browman and Goldstein (1987), we expect to be able to model the schwa effects we have observed as resulting from a structure in which there is an active gesture for the medial schwa, but complete temporal overlap of this gesture and the gesture for the following vowel. The blending caused by this overlap should yield the V2 effect on schwa, while the V1 effects should emerge as a passive consequence of the differing initial conditions for movements out of different preceding vowels.

An example of this type of organization is shown in Figure 12, which is the gestural score we hypothesized for the utterance /piypax'paapax/.

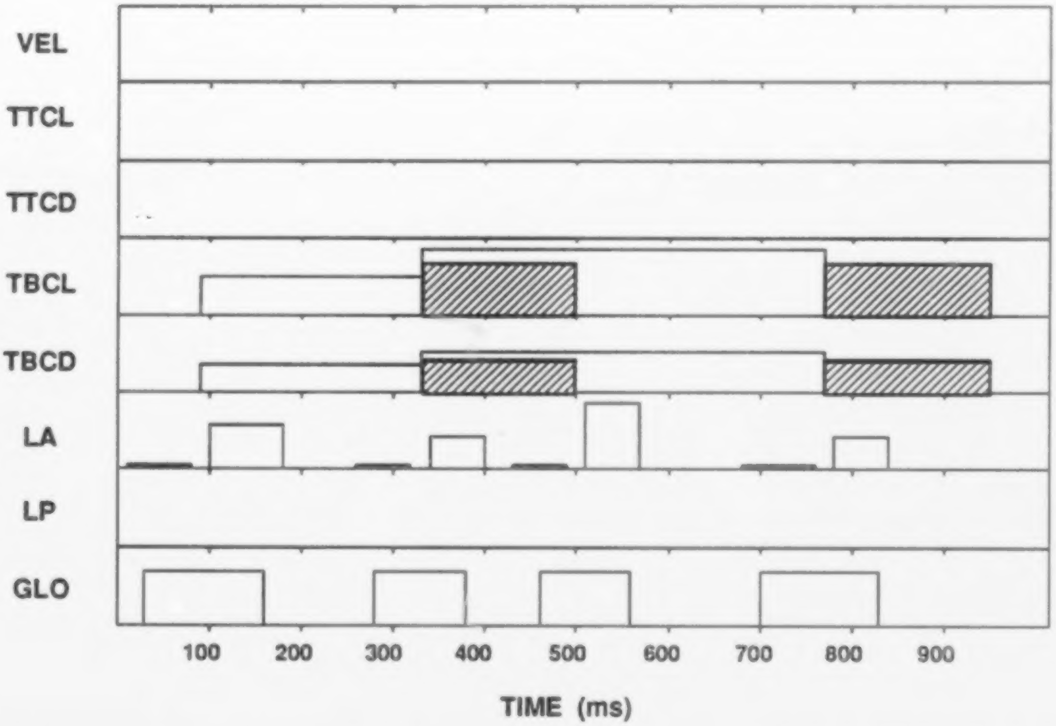


Figure 12. **Gestural score for /piypax'paapax/. Tract variable channels displayed, from top to bottom, are: VELum, Tongue Tip Constriction Location and Constriction Degree, Tongue Body Constriction Location and Constriction Degree, Lip Aperture, Lip Protrusion, and GLOttis. Boxes indicate gestural activation; the shaded boxes indicate activation for schwa.**

As in Figure 3, each box indicates the activation interval of a particular gestural control regime, that is, an interval of time during which the behavior of the particular tract variable is controlled by a second order dynamical system with a fixed "target" (equilibrium position), frequency, and damping. The height of the box represents the tract variable "target." Four lip aperture (LA) closure-and-release gestures are shown, corresponding to the four consonants. The closure and release components of these gestures are shown as separate boxes, with the closure components having the smaller target for LA, i.e., smaller interlip distance. In addition, four tongue body gestures are shown, one for each of the vowels—V1, schwa, V2, schwa. Each of these gestures involves simultaneous activation of two tongue body tract variables, one for constriction location (TBCL) and one for constriction degree (TBCD). The control regimes for the V1 and medial schwa gestures are contiguous and non-overlapping, whereas the V2 gesture begins at the same point as the medial schwa and thus completely overlaps it. In other words, during the acoustic realization of the schwa (approximately), the schwa and V2 gestural control regimes both control the tongue movements; the schwa relinquishes active control during the following consonant, leaving only the V2 tongue gesture active in the next syllable. While the postulation of an explicit schwa gesture overlapped by V2 was motivated by the particular results of section 2, the general layout of gestures in these utterances (their durations and overlap) was based on stiffness and phasing principles embodied in the linguistic model (Browman & Goldstein, 1987).

Gestural scores for each of the twenty-five utterances were produced. The activation intervals were identical in all cases; the scores differed only in the TBCL and TBCD target parameters for the different vowels. Targets used for the full vowels were those in our tract variable dictionary. For the schwa, the target values (for TBCL and TBCD) were calculated as the mean of the targets for the five full vowels. The gestural scores were input to the task dynamic model (Saltzman, 1986), producing motions of the model articulators of the articulatory synthesizer (see Figure 1). For example, for utterance /piypax'piypax/, Figure 13 shows the resulting motions (with respect to a fixed reference on the head) of two of the articulators—the center of the tongue body circle (C), and the lower lip, superimposed on the gestural score. Motion of the tongue is shown in both horizontal and vertical dimensions, while only vertical motion of the lower lip is shown. Note that the lower lip moves up for lip closure (during the regimes with the small LA value). Figure 14 shows the results for /piypax'paapax/.

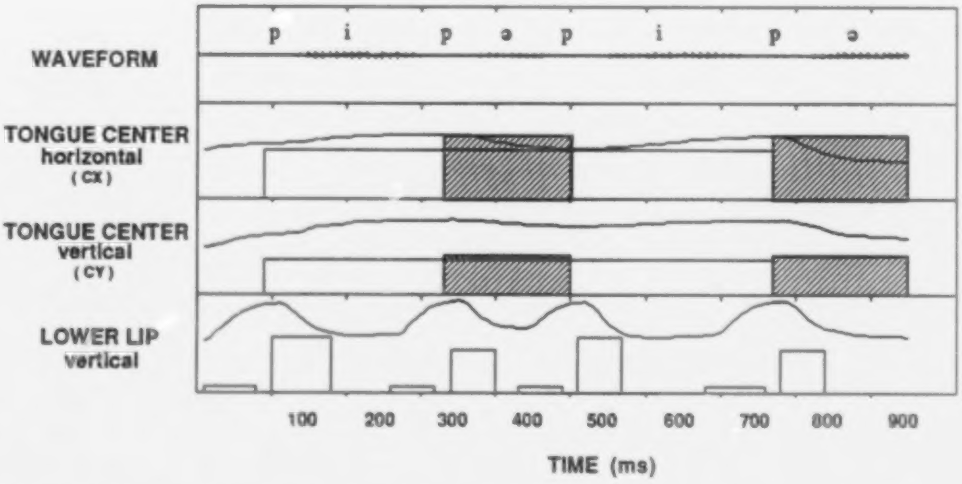


Figure 13. Gestural score for /piypax'piypax/ (IPA symbols are used in the figure). Generated movements (curves) are shown for the tongue center and lower lip. Boxes indicate gestural activation; the shaded boxes indicate activation for schwa.

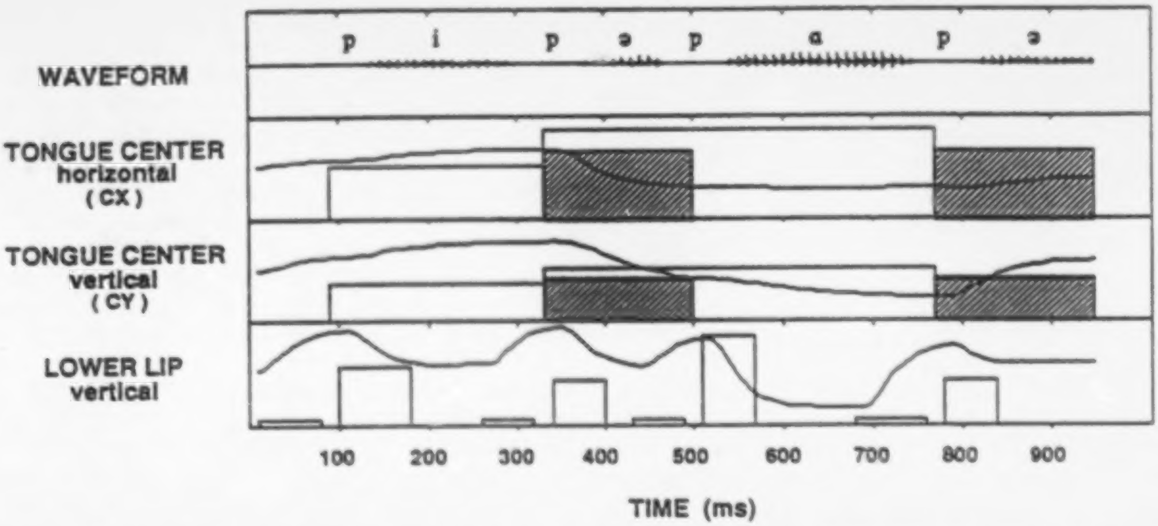


Figure 14. Gestural score for /piypax'paapax/ (IPA symbols are used in the figure). Generated movements (curves) are shown for the tongue center and lower lip. Boxes indicate gestural activation; the shaded boxes indicate activation for schwa.

The articulator motions in the simulations can be compared to those of the data in the previous section (Figures 4 and 5). One difference between the model and the data stems from the fact that the major portion of the tongue dorsum is modeled as an arc of circle, and therefore all points on this part of the dorsum move together. Thus, it is not possible to model the differential patterns of motion exhibited by the middle (M) and rear (R) of the dorsum in the X-ray data. In general, the motion of CX is qualitatively similar to both MX and RX (which, recall, are highly correlated). For example, both the data and the simulation show a small backward movement for the schwa in /piypax'piypax/; in /piypax'paapax/, both show a larger backwards movement for schwa, with the target for /aa/ reached shortly thereafter, early in the acoustic realization of V2. The motion of CY in the simulations tends to be similar to that of MY in the data. For example, in /piypax'paapax/, CY moves down from /iy/ to schwa to /aa/, and the target for /aa/ tends to be achieved relatively late, compared to CX. Movements corresponding to RY motions are not found in the displacement of the tongue body circle, but would probably be reflected by a point on the part of the model tongue's surface that is further back than that section lying on the arc of a circle.

The model articulator motions were analyzed in the same manner as the X-ray data, once the time points for measurement were determined. Since for the X-ray data we assumed that displacement extrema indicated the effective target for the gesture, we chose the effective targets in the simulated data as the points to measure. Thus, points during V1 and V2 were chosen that corresponded to the point at which the vowel gestures (approximately) reached their targets and were turned off (right hand edges of the tongue boxes in Figures 13 and 14). For schwa, the "tongue" reference point was chosen at the point where the schwa gesture was turned off, while the "lip" reference was chosen at the lowest point of the lip during schwa (the same criterion as for the X-ray data).

The distribution of the model full vowels in the mid-sagittal plane (CX × CY) is shown in Figure 15. Since the vowel gestures are turned off only after they come very close to their targets, there is very little variation across the 10 tokens of each vowel. The distribution of schwa at the "tongue" reference point is shown in Figure 16, labelled by the identity of V2 (in a) and V1 (in b), with the full vowel ellipses added for comparison. At this reference point that occurs relatively late, the vowels are clustered almost completely by V2, and the tongue center has moved a substantial portion of the way toward the following full vowel. The distribution of schwa values at the "lip" reference point is shown in Figure 17, labelled by the identity of V2 (in a), and of V1 (in b). Comparing Figure 17(a) with Figure 16(a), we can see that there is considerably more scatter at the "lip" point than at the later "tongue" point.

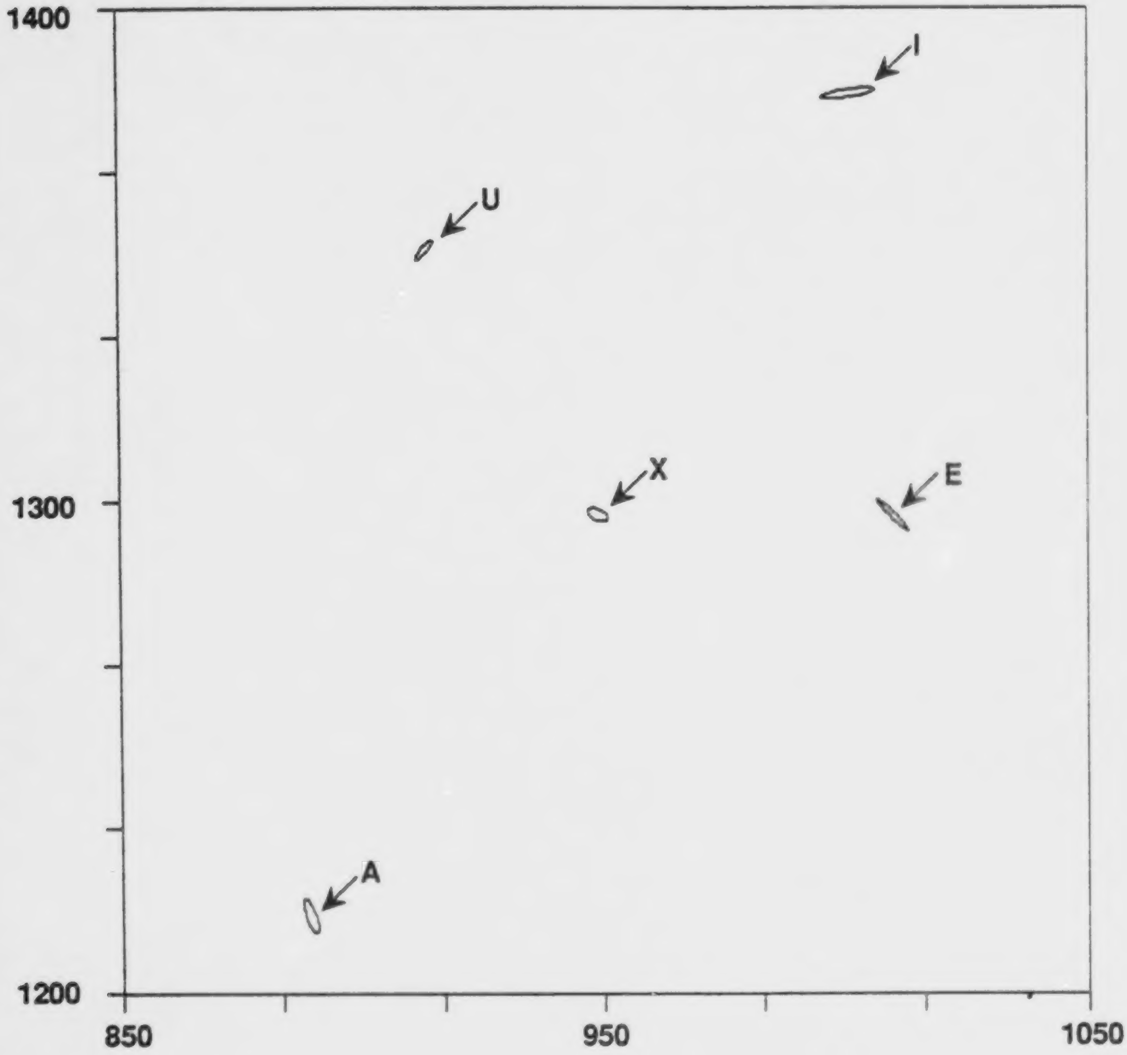


Figure 15. **Tongue center (C) positions for model full vowels, displayed in mid-sagittal plane with head facing to the right. The ellipses indicate two standard deviations along axes determined by principal component analysis. Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are ASY units (= .09 mm).**

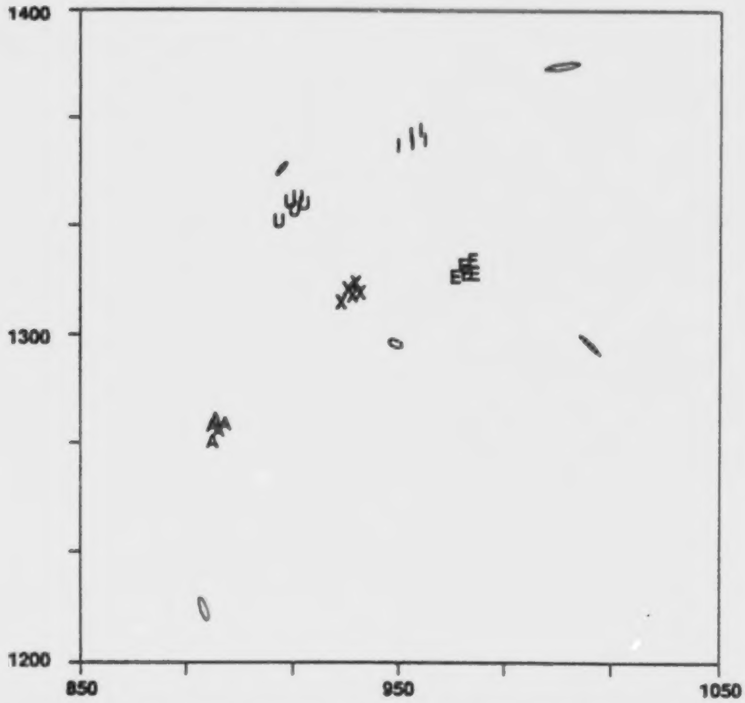
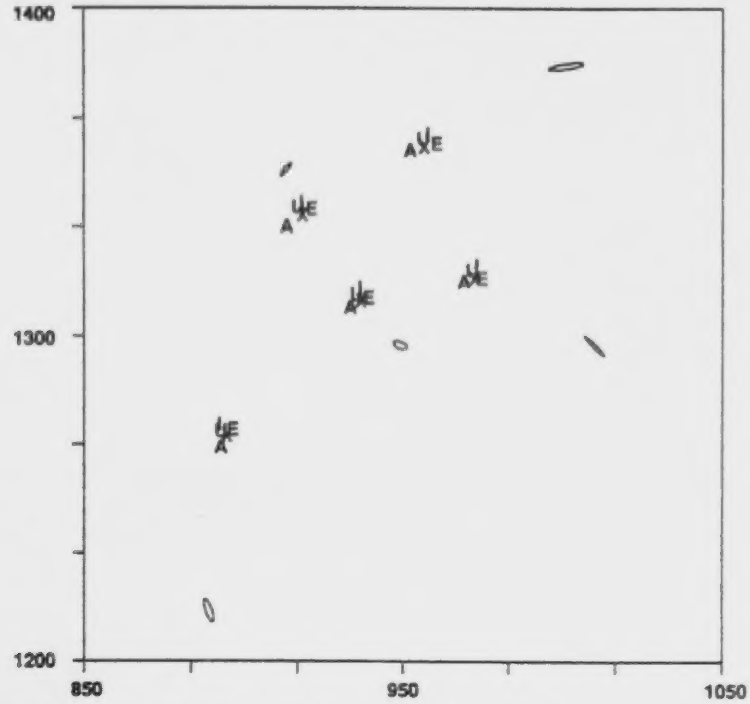


Figure 16. **Tongue center (C) positions for model schwa at "tongue" reference points, displayed in right-facing midsagittal plane as in Figure 15. The ellipses are from the model full vowels (Figure 15), for comparison. Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are ASY units (= .09 mm).** *(a)* **Model schwa positions labelled by the identity of the following vowel (V2).**



(b) **Model schwa positions labelled by the identity of the preceding vowel (V1).**

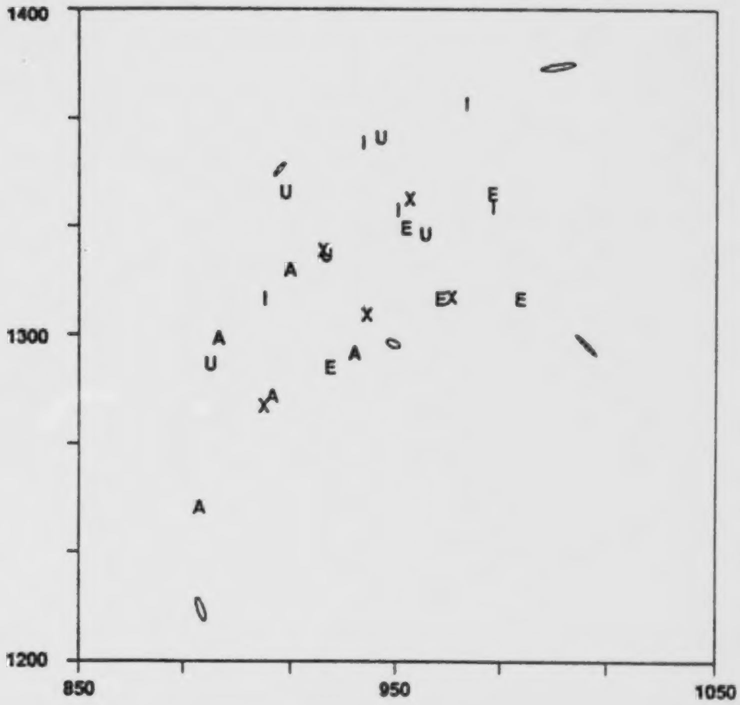
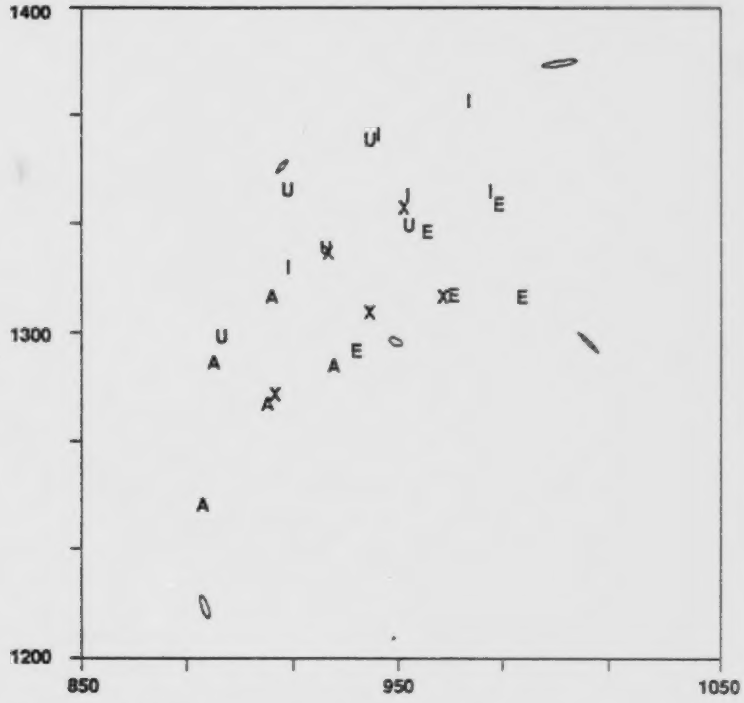


Figure 17. **Tongue center (C) positions for model schwa at "lip" reference points, displayed as in Figure 16 (including ellipses from Figure 15). Symbols I = ARPAbet /iy/, U = /uw/, E = /eh/, X = /ah/, and A = /aa/. Units are ASY units (= .09 mm).** ***(a)*** **Model schwa positions labelled by the identity of the following vowel (V2).**



(b) **Model schwa positions labelled by the identity of the preceding vowel (V1).**

We tested whether the simulations captured the regularities of the X-ray data by running the same set of regression analyses on the simulations as were performed on the X-ray data. The results are shown in Table 2, which has the same format as the X-ray data results in Table 1.

Table 2. *Regression results of simulations.*

	"lip" reference point			"tongue" reference point	
	Terms	*Standard Error*		*Terms*	*Standard Error*
CX	k+v1+v2	7.5	CX	k+v2+v1	5.2
	v1+v2	9.14		k+v2	5.6
	k+v1	20.4		v2+v1	13.2
	k	29.2		v2	20.0
	v1	33.3		k	28.6
CY	k+v1+v2	4.7	CY	k+v2+v1	3.2
	v1+v2	13.5		k+v2	3.9
	k+v1	19.7		v2+v1	18.8
	k	29.1		v2	28.0
	v1	41.6		k	30.4

The same patterns are found for the simulations as for the data. At the "tongue" reference point, for both CX and CY the best two-term prediction involves the schwa component (constant) and V2, and this prediction is as good as that using all three terms. Recall that this was the case for all pellet dimensions except for MY, whose differences were attributed to differences in the time point at which this dimension was measured. (In the simulations, CX and CY were always measured at the same point in time.) These results can be seen graphically in Figure 18, where the top row of panels shows the relation between V1 and schwa, and the bottom row shows the relation between V2 and schwa. The same systematic relation between schwa and V2 can be seen in the bottom row as in Figure 10 or the X-ray data, that is, no crossover. (The lack of systematic relations between V1 and schwa in the X-ray data, indicated by the crossovers in the top row of Figure 10, is captured in the simulations in Figure 18 by the lack of variation for the schwa in the top row.) Thus, the simulations capture the major relation between the schwa and the surrounding full vowels at the "tongue" reference point.

At the earlier "lip" reference point, the simulations also capture the patterns shown by the data. For both CX and CY, the three-term predictions in Table 2 show substantially less error than the best two-term prediction. This was also the case for the data in Table 1 (except for RX), where V1, V2 and a schwa component (constant) all contributed to the prediction of the value during schwa. This can also be seen in the graphs in Figure 19, which shows a systematic relationship with schwa for both V1 and V2.

In summary, for the simulations, just as for the X-ray data, V1 contributed to the pellet position at the "lip" reference, but not to the pellet position at the "tongue" point, while V2 and an independent schwa component contributed at both points. Thus, our hypothesized gestural structure accounts for the regularities observed in the data. The gestural control regime for V2 begins simultaneously with that for schwa and overlaps it throughout its active interval. This accounts for the fact that V2 and schwa effects can be observed throughout the schwa, as both gestures unfold together. However, V1 effects are passive consequences of the initial conditions when the schwa and V2 gestures are "turned on," and thus, their effects disappear as as the tongue position is attracted to the "target" (equilibrium position) associated with the schwa and V2 regimes.

3.1 Other simulations

While the X-ray data from the subject analyzed here argue against the strongest form of the hypothesis that schwa has no tongue target, we decided nevertheless to perform two sets of simulations incorporating the strong form of an "unspecified" schwa to see exactly where and how they would fail to reproduce the subject's data. In addition, if the synthesized speech were found to be correctly perceived by listeners, it would suggest that this gestural organization is at least a possible one for these utterances, and might be found for some speakers. In the first set of simulations, one of which is exemplified in Figure 20, the gestural scores took the same form as in Figure 12, except that the schwa tongue body gestures were removed. Thus, active control of V2 began at the end of V1, and, without a schwa gesture, the tongue trajectory moved directly from V1 to V2. During the acoustic interval corresponding to schwa, the tongue moved along this V1-V2 trajectory. The resulting simulations in most cases showed a good qualitative fit to the data, and produced utterances whose medial vowels were perceived as schwas. The problems arose in utterances in which V1 and V2 were the same (particularly when they were high vowels).

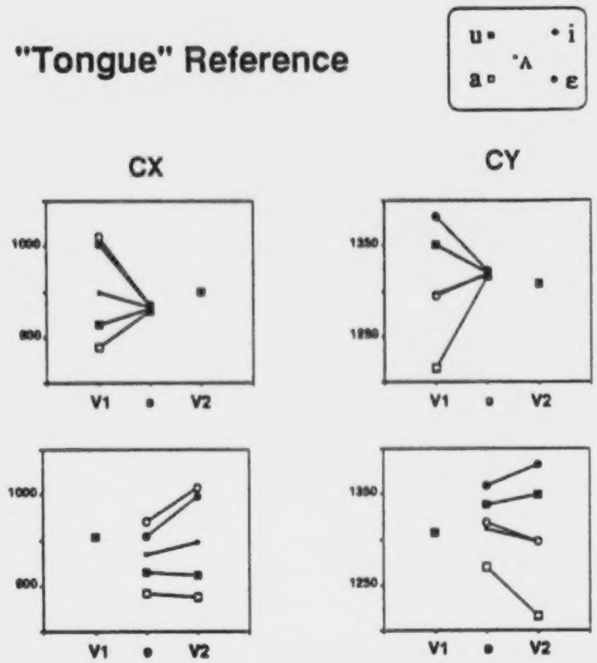


Figure 18. **Relation between model full vowel tongue center positions and tongue center positions at "tongue" reference point for model schwas. The top row displays the tongue center positions for utterances with the indicated initial vowels, averaged across 5 utterances (each with a different final vowel). The bottom row displays the averaged tongue center positions for utterances with the indicated final vowels. Units are ASY units (= .09 mm).**

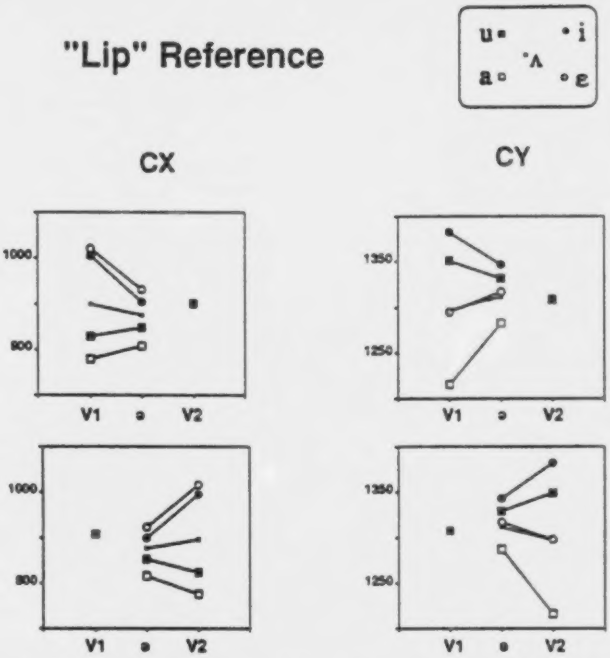


Figure 19. **Relation between model full vowel tongue center positions and tongue center positions at "lip" reference point for model schwas. The top row displays the tongue center positions for utterances with the indicated initial vowels, averaged across 5 utterances (each with a different final vowel). The bottom row displays the averaged tongue center positions for utterances with the indicated final vowels. Units are ASY units (= .09 mm).**

Figure 20 portrays the simulation for /piypax'piypax/: the motion variables generated can be compared with the data in Figure 4. The "dip" between V1 and V2 was not produced in the simulation, and, in addition, the medial vowel sounded like /iy/ rather than schwa. This organization does not, then, seem possible for utterances where both V1 and V2 are high vowels.

We investigated the worst utterance (/piypax'piypax/) from the above set of simulations further, generating a shorter acoustic interval for the second vowel (the putative schwa) by decreasing the interval (relative phasing) between the bilabial gestures on either side of it. An example of a score with the bilabial closure gestures closer together is shown in Figure 21. At relatively short durations as in the figure (roughly 50 ms), the percept of the second vowel changed from /iy/ to schwa. Thus, the completely targetless organization may be workable in cases where the surrounding consonants are only slightly separated. In fact, this suggests a possible historical source for epenthetic schwa vowels that break up heterosyllabic clusters. They could arise from speakers increasing the distance between the cluster consonants slightly, until they no longer overlap. At that point, our simulations suggest that the resulting structure would be perceived as including a schwa-like vowel.

The second set of simulations involving an "unspecified" schwa used the same gestural organization as that portrayed in the score in Figure 20, except that the V2 gesture was delayed so that it did not begin right at the offset of the V1 gesture. Rather, the V2 regime began approximately at the beginning of the third bilabial closure gesture, as in Figure 22. Thus, there was an interval of time during which no tongue body gesture was active. The motion of the tongue tract variables during this interval was determined by the neutral positions associated with the specific tongue body articulators, and by the motion of the jaw, which was implicated in the ongoing bilabial closure and release gestures. The results, displayed in Figure 22, showed that the previous problem with /piypax'piypax/ was solved, since during the unspecified interval between the two full vowels, the tongue body lowered (from /iy/ position) and produced a perceptible schwa. Unfortunately, this "dip" between V1 and V2 was seen for all combinations of V1 and V2, which was not the case in the X-ray data. For example, this dip can be seen for /paapax'paapax/ in Figure 23; in the X-ray data, however, the tongue raised slightly during the schwa, rather than lowering. (The "dip" occurred in all the simulations because the neutral position contributing to the tongue body movement was that of the tongue body articulators rather than that of the tongue body tract variables, and thus was relative to the jaw, which, in turn, was lowering as part of the labial release). In addition, because the onset for V2 was so late, it would not be possible for V2 to affect the schwa at the "lip" reference point, as was observed in the X-ray data. Thus, this hypothesis also failed to capture important aspects of the data. The best hypothesis remains the one tested first—where schwa has a target of sorts, but is still "colorless," in that its target is the mean of all (flanking) vowels, and is completely overlapped by the following vowel.

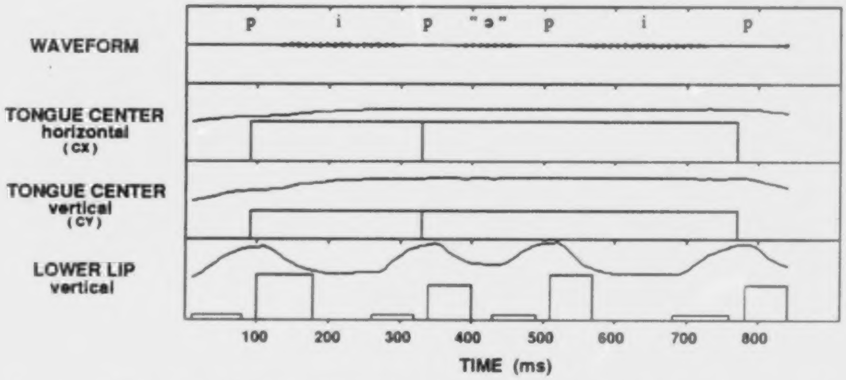


Figure 20. Gestural score plus generated movements for /piyp_'piyp_/, with no activations for schwa. The acoustic interval between the second and third bilabial gestures is perceived as an /iy/. (IPA symbols are used in the figure).

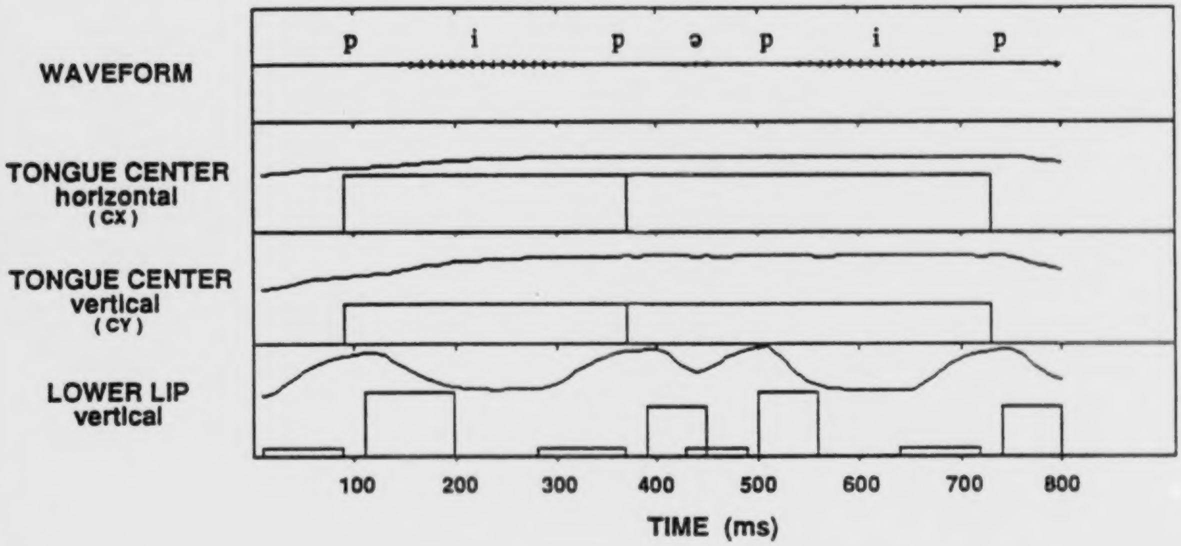


Figure 21. **The same gestural score for /piyp_'piyp_/ as in Figure 20, but with the second and third bilabial gestures closer together than in Figure 20. The acoustic interval between the second and third bilabial gestures is perceived as a schwa. Generated movements (curves) are shown for the tongue center and lower lip. (IPA symbols are used in the figure).**

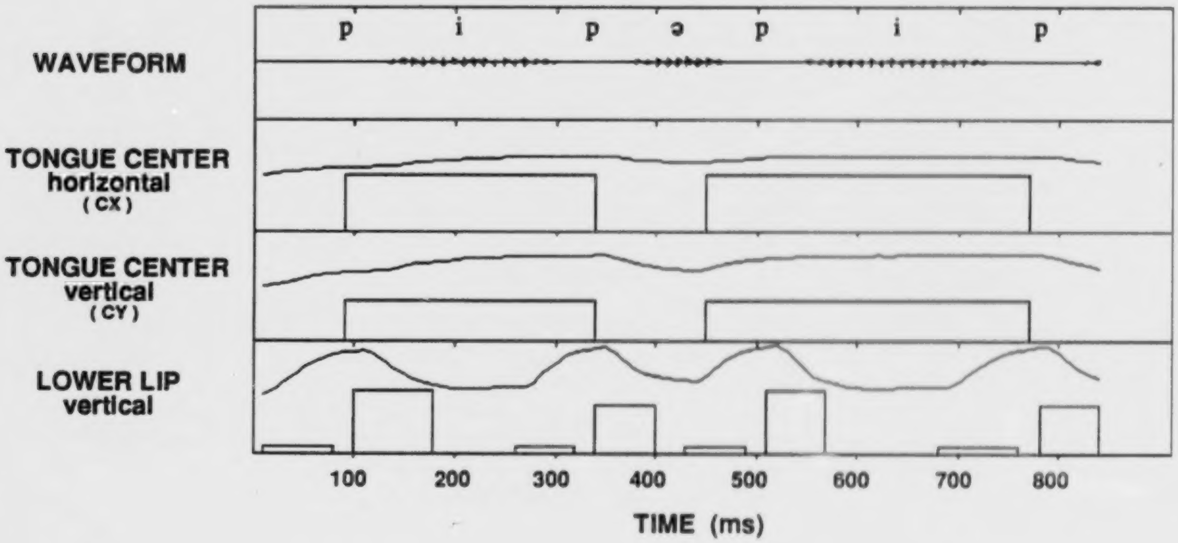


Figure 22. **The same gestural score for /piyp_'piyp_/ as in Figure 20, but with the onset of the second full vowel /iy/ delayed. The acoustic interval between the second and third bilabial gestures is perceived as a schwa. Generated movements (curves) are shown for the tongue center and lower lip. (IPA symbols are used in the figure).**

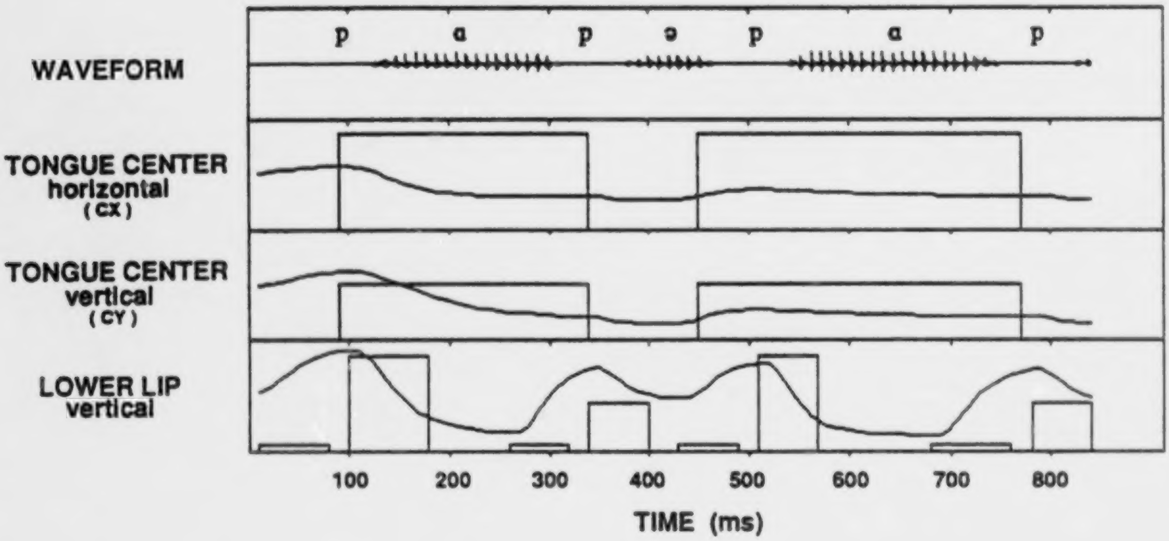


Figure 23. The same gestural score as in Figure 22, except with tongue targets appropriate for the utterance /paap_'paap_/. The acoustic interval between the second and third bilabial gestures is perceived as a schwa. Generated movements (curves) are shown for the tongue center and lower lip. (IPA symbols are used in the figure).

4. CONCLUSION

We have demonstrated how an explicit gestural model of phonetic structure, embodying the possibilities of underspecification ("targetlessness") and temporal overlap ("co-production"), can be used to investigate the contextual variation of phonetic units, such as schwa, in speech. For the particular speaker and utterances that we analyzed, there was clearly some warping of the V1-V2 trajectory towards a neutral position for an intervening schwa. The analyses showed that this neutral position has to be defined in the space of tract variables (the linguistically-relevant goal space), rather than being the consequence of neutral positions for individual articulators. Therefore, a target position for schwa was specified, although this target is completely predictable from the rest of the system; it corresponds to the mean tongue tract variable position for all the full vowels.

The temporally overlapping structure of the gestural score played a key role in accounting for the time course of V1 and V2 effects on schwa. These effects were well modeled by a gestural score in which active control for schwa was completely overlapped by that for V2. This overlap gave rise to the observed anticipatory effects, while the carryover effects were passive consequences of the initial conditions of the articulators when schwa and V2 begin. (This fits well with studies that have shown qualitative asymmetries in the nature of carryover and anticipatory effects, see Recasens, 1987).

How well the details of the gestural score will generalize to other speakers and other prosodic contexts remains to be investigated. There is known to be much individual variation in the strength of anticipatory vs. carryover coarticulation in utterances like those employed here, and also in the effect of stress (Fowler, 1981b; Magen, 1989). In addition, schwas with different phonological/morphological characteristics (e.g. in plural /-axz/ and past tense /-axd/) may show different behavior, either with respect to overlap or targetlessness. The kind of modeling developed here provides a way of analyzing the complex quantitative data of articulation so that phonological issues such as these can be addressed.

REFERENCES

Alfonso, P., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech, 25,* 151-173.

Anderson, S. (1982). The analysis of French schwa. *Language, 58,* 535-573.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica*, 38, 9-20.

Browman, C.P., & Goldstein, L. (1987). Tiers in Articulatory Phonology, with some implications for casual speech. Haskins Laboratories *Status Report on Speech Research*, *SR-92*, 1-30. To appear in J. Kingston & M.E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge: Cambridge University Press.

Browman, C. P., Goldstein, L., Saltzman, E., & Smith, C. (1986). GEST: A computational model for speech production using dynamically defined articulatory gestures. *Journal of the Acoustical Society of America*, *A 80*. S97. (Abstract)

Fowler, C. A. (1977). *Timing control in speech production*. Bloomington: Indiana University Linguistics Club.

Fowler, C. A. (1981a). Perception and production of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, *24*, 127-139.

Fowler, C. A. (1981b). A relationship between coarticulation and compensatory shortening. *Phonetica*, *38*, 35-50.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, *62*, 693-707.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, *5*, 275-292.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language*, *56*, 485-502.

Ladefoged, P. (1982). *A course in phonetics* (2nd ed.). New York: Harcourt Brace Jovanovich.

Ladefoged, P., & Lindau, M. (1989). Modeling articulatory-acoustics relations: A comment on Stevens' "On the quantal nature of speech." *Journal of Phonetics*, *17*, 99-106.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.

Magen, H. (1989). *An acoustic study of vowel-to-vowel coarticulation in English*. Doctoral dissertation, Yale University, Department of Linguistics.

Miller, J. E., & Fujimura, O. (1982). Graphic displays of combined presentations of acoustic and articulatory information. *The Bell System Technical Journal*, *61*, 799-810.

Neary, T. M. (1980). On the physical interpretation of vowel quality: Cinefluorographic and acoustic evidence. *Journal of Phonetics*, 8, 213-241.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151-168.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. (Research Monograph No. 53). Cambridge, MA: The MIT Press.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, *15*, 299-312.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, *70*, 321-328.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns (Experimental Brain Research Series 15)*. New York: Springer-Verlag.

Saltzman, E., Goldstein, L., Browman, C. P., & Rubin, P. E. (1988). Dynamics of gestural blending during speech production. Presented at First Annual International Neural Network Society (INNS), Boston, September 6-10, 1988.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review*, *94*, 84-106.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* *1(4)*, 333-382..

Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Papers*, 23. Lund University.

FOOTNOTES

[*] Paper presented at Second Conference on Laboratory Phonology, Edinburgh, Scotland, 28 June 1989-3 July 1989.

[†] Also Department of Linguistics, Yale University.

[1] ARPAbet - IPA equivalents:

ARPAbet	IPA
iy	i
eh	ɛ
aa	a
ah	ʌ
uw	u
ax	ə

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 220-230

Methods for Least-squares Parameter Identification for Articulatory Movement and the Program PARFIT

Richard S. McGowan, Caroline L. Smith, Catherine P. Browman, and Bruce A. Kay[†]

The method of parameter identification using nonlinear least-squares curve fitting is discussed in this note. After the algorithm (a multidimensional Newton's method) is described in general, some of our particular experiences in fitting damped sinusoids to simulated data are discussed. We find that it is possible to identify parameters for fitting the simulated data by using constraints to eliminate some parameters and to limit the range of others.

INTRODUCTION

The problem of estimating the parameter values of a dynamical system characterizing articulatory movement is addressed in this note. It is supposed that the trajectory data of the articulators as a function of time may be fit by a model function of time. This function is the solution to a differential equation describing a dynamical system, so that the parameters of the model function correspond to the parameters of the dynamical system. The model function here is chosen to be that of a damped sinusoid with a constant level added, the solution of the differential equation describing a linear damped mass-spring system. The criterion for fitting the model function to the data is that of least-squares error, and the parameters of the model equation, such as frequency and damping ratio, are varied to achieve this minimum least-squares criterion. This fitting problem belongs to the general class of nonlinear least-squares fitting problems, because the model function depends nonlinearly on the parameters.

In the first part of this note, some nonlinear least-squares fitting algorithms are discussed (nonlinear regression). In the the second part of this note, the particular procedure devised for fitting articulatory data to damped sinusoidal curves, the program PARFIT, is discussed.

THEORY AND ALGORITHM: NONLINEAR LEAST SQUARES

Given position data as a function of time $y^d(t)$, it will be supposed that the data can be digitized to obtain samples Δt apart, so that position at time $t_n = n\Delta t$ is given by $y^d{}_n$. Suppose that a model function is given, $y(t, \mathbf{a})$, where $\mathbf{a}$ is a vector of parameters, $\mathbf{a} = (a_1, a_2, \ldots, a_N)$. The solution function for a damped sinusoidal model is

$$y(t,a) = e^{-\alpha t}\{A\cos(\omega t) + B\sin(\omega t)\} + C \qquad (1)$$

where $\mathbf{a} = (\alpha, A, B, \omega, C)$. This function is the solution to the differential equation

$$\ddot{y} + \beta\dot{y} + \omega_0{}^2(y - C) = 0 \qquad (2)$$

The relationships between the coefficients in the differential equation and the parameters in the solution are as follows. $\alpha = \beta/2$ is the damping

---

The authors gratefully acknowledge the support of NIH grants HD-1994 and NS-13617 and NSF grant BNS-8520709 to Haskins Laboratories. Many helpful comments were made by Elliot Saltzman.

factor, ω_0 is the natural circular frequency, and ω is the observed circular frequency, where $\omega = (\omega_0^2 - \alpha^2)^{1/2}$. C is the constant, or D.C., level. The parameters A and B must be determined by other constraints, such as initial position and velocity. The damping ratio, a quantity to be used later, is defined as, α / ω_0 (Braun, 1983).

For least-squares fitting the chi-square function to be minimized is defined as

$$\chi^2(\mathbf{a}) = \sum_{n=1}^{M} \frac{(y_n^d - y(t_n, \mathbf{a}))^2}{\sigma_n^2} \qquad (3)$$

where σ_n is the expected standard deviation in the data at time t_n and M is the number of data points. The data trajectory, y^d, is presumed to be from an ensemble of trajectories, where the data points, $y^d{}_n$ are independent, random variables. In the case where these random variables are normally distributed, the chi-square function is a random variable with a chi-square distribution. In the following it is assumed that all the data points are measured with the same certainty, that is, they have the same expected standard deviation, although this quantity may have an unknown value. The chi-square function is a scalar-valued function that provides a distance metric between a single data trajectory and the model function. The problem of finding the best fit between this single trajectory and the model function is the focus of the discussion here, although statistical questions will be addressed briefly later in the paper.

The problem can be pictured in three dimensions if there are two components in the parameter vector of the model function. For instance, take the model function to be A cos(ωt), with variable parameters A and ω. The chi-square function can be represented locally in the region of the minimum as a surface over the A-ω plane (see Figure 1). The surface is supposed smooth in the region of the minimum of chi-square attained at A^* and ω^*. Note that the curvature is upward in all directions at the minimum.

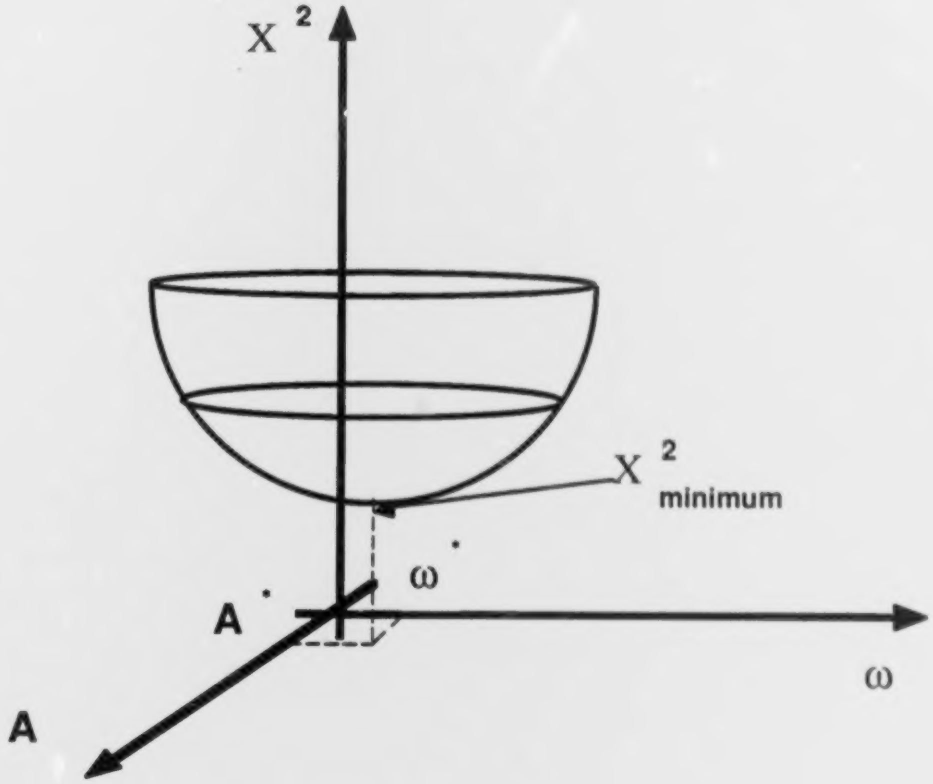


Figure 1. Chi-square surface.

Least-squares algorithms are meant to find a global minimum to the chi-square function. That is, they are supposed to find a point $\mathbf{a}^*$, such that $\chi^2(\mathbf{a}) >= \chi^2(\mathbf{a}^*)$ for all $\mathbf{a}$. Initially, the algorithm for finding a minimum of the chi-square function given by Press, Flannery, Teukolsky, and Vetterling (1986) is explained. Later, variations to this basic algorithm are introduced.

Suppose y, and hence χ^2, to be a differentiable function of the parameter vector $\mathbf{a}$. A necessary condition that χ^2 attain a minimum at $\mathbf{a}^*$ is that the ***gradient*** of χ^2 be zero at $\mathbf{a}^*$.

$$\left.\frac{\partial\chi^2}{\partial a_k}\right|_{\mathbf{a}^*} = 0. \tag{4}$$

for k = 1,...,N.
In vector notation

$$\nabla\chi^2\big|_{\mathbf{a}^*} = 0. \tag{5}$$

Assuming that χ^2 is three times differentiable with respect to $\mathbf{a}$, the gradient of χ^2 can be approximated near $\mathbf{a}$ by a Taylor series expansion

$$\nabla\chi^2\big|_{\mathbf{a}^*} = \nabla\chi^2\big|_{\mathbf{a}} + \mathbf{D}\cdot(\mathbf{a}^* - \mathbf{a}) + \vartheta|\mathbf{a}^* - \mathbf{a}|^2 \tag{6}$$

where

$$\mathbf{D} = [D_{lk}] = \left[\frac{\partial^2\chi^2}{\partial a_l\,\partial a_k}\right]$$

for k, l = 1,..., N . **D** is known as the *Hessian matrix*, and the final term denotes the asymptotic order of the error in the approximation.

Neglecting the error term, and using equation (5), equation (6) can be solved for $\mathbf{a}^*$

$$\mathbf{a}^* = \mathbf{a} - \mathbf{D}^{-1}\cdot\nabla\chi^2\big|_{\mathbf{a}} \tag{7}$$

(Of course, **D** must be nonsingular for this to be valid.) The second term on the right in equation (7) provides a multidimensional correction to $\mathbf{a}$ for finding the vector root of the gradient of χ^2. This equation, used iteratively, constitutes ***Newton's method***.[1]

Newton's method, as given in equation (7) is not the only procedure that may be used in searching for the root $\mathbf{a}^*$. In fact, it may not work as well as other procedures when $\mathbf{a}$ is far away from $\mathbf{a}^*$, where a ***steepest descents*** algorithm may work better. The relationship between Newton's method and steepest descents is explored a little later in this paper. This relationship is best illustrated after equation (7) has been reinterpreted in the light of linear least-squares algorithms. This reinterpretation will also lead to considerations on how to compute the correction term in equation (7), which is the most important numerical consideration here.

It should be noted that solving equation (5) does not assure that a local minimum has been found, much less a global minimum. To assure the solution to equation (5), $\mathbf{a}^*$, is a local minimum, the matrix **D** should be checked for positive definiteness at $\mathbf{a}^*$, so that we can be assured of upward curvature of the χ^2 surface at the point $\mathbf{a}^*$ in all directions. This can be done numerically, but in practice only the smallness of χ^2 is checked. Also, this appears to be good enough for our purposes, but does not guarantee that a minimum point has been attained. There is no sure way to know whether a local minimum is also a global minimum, but by starting the iteration from different initial parameters and choosing the iteration with the least error can give us confidence that a global minimum has been attained.

Turning now to the computation of **D** and $\mathbf{D}^{-1}$, note that

$$D_{lk} \equiv \frac{\partial^2\chi^2}{\partial a_l\,\partial a_k} = 2\sum_{n=1}^{M}\left[\frac{\partial y(t_n,\mathbf{a})}{\partial a_l}\frac{\partial y(t_n,\mathbf{a})}{\partial a_k} - \left(y_n^d - y(t_n,\mathbf{a})\right)\frac{\partial^2 y(t_n,\mathbf{a})}{\partial a_l\,\partial a_k}\right] \tag{8}$$

Press et al. (1986) argue that the second term in the sum should be small because the errors, $y^d{}_n - y(t_n,\mathbf{a})$, should be uncorrelated with the model. Therefore equation (8) can be approximated as

$$D_{lk} \approx 2\sum_{n=1}^{M}\frac{\partial y(t_n,\mathbf{a})}{\partial a_l}\frac{\partial y(t_n,\mathbf{a})}{\partial a_k} \tag{9}$$

This approximation is particularly useful, because it means that the matrix **D** can be factored. Let

$$\mathbf{F} = \begin{pmatrix} \frac{\partial y(t_1)}{\partial a_1} & \cdots & \frac{\partial y(t_1)}{\partial a_N} \\ \vdots & & \vdots \\ \frac{\partial y(t_M)}{\partial a_1} & \cdots & \frac{\partial y(t_M)}{\partial a_N} \end{pmatrix} \tag{10}$$

or

$$\mathbf{F} = [F_{ij}] = \left[\frac{\partial y(t_i)}{\partial a_j}\right]$$

Then, from the definition of **D** and equations (9) and (10)

$$\mathbf{D} \approx 2\,\mathbf{F}^T\mathbf{F} \tag{11}$$

where T denotes matrix transpose.

Another useful substitution can be made using the matrix **F**.

Let $\delta y = [\delta y_n] = [y^d_n - y(t_n)]$. From equation (3)

$$\nabla \chi^2 = -2\mathbf{F}^T \cdot \delta\, y \tag{12}$$

Using equations (11) and (12), equation (7) can be written

$$\delta a = \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T \cdot \delta\, y \tag{13}$$

where $\delta \mathbf{a} = \mathbf{a}^* - \mathbf{a}$.

Recall that **D** is assumed to be nonsingular. In this case $(\mathbf{F}^T\mathbf{F})^{-1}\,\mathbf{F}^T$ is recognized as the pseudoinverse of **F**, $\mathbf{F}^I$, which is well known in algorithms solving linear least-squares problems (Stewart, 1973). With the iteration for Newton's method in the form of equation (13) a connection between Newton's method and linear least squares can be made. (This is not surprising, because Newton's method amounts to finding a linear approximation to the function for which the root is being found.) To make this connection an alternative method for deriving equation (13) is sought. Simultaneously, this points toward numerical methods, which are, perhaps, superior to the direct inversion of **D**.

Given that the correct model has been chosen, the discrepancy between the data and the model is given by

$$y^d - y|_{\mathbf{a}} = \mathbf{F}|_{\mathbf{a}} \cdot \delta\, a + \vartheta\left|\delta a\right|^2 + \varepsilon \tag{14}$$

where y^d is the vector of data points, $y|_a$ is the vector of model function values evaluated with the parameter vector **a**, and ε is a vector of random noise components. If **a** is the vector of parameters giving the best fit, the first and second terms on the right-hand-side disappear, and the difference between the model and data is the random noise, ε. To the same approximation made in equations (7) and (11) the final two terms in equation (14) can be neglected to obtain

$$\delta y = \mathbf{F}|_{\mathbf{a}} \cdot\ \delta a \tag{15}$$

In the following, it is supposed that there are at least as many data points as parameters, so that M >= N. (In fact, there should be several more data points than parameters to have confidence in the fit.) Equation 15 may not have a solution, but we can ask to minimize the distance between δy and $\mathbf{F}\,\delta\,\mathbf{a}$.

$$\|\,\delta y\ - \mathbf{F} \cdot \delta\ a\,\|^2 = \sum_{n=1}^{M} \left(\delta y_n - \left(\mathbf{F} \cdot \delta\ a\right)_n\right)^2 \tag{16}$$

where $\|\ \|$ denotes the Euclidian vector norm.

This is a linear least-squares problem, the solution to which is given by

$$\delta a = \mathbf{F}^I \cdot\ \delta y \tag{17}$$

where $\mathbf{F}^I$ is the pseudoinverse of **F**. In the case that **F** has linearly independent columns, the solution to the minimization problem is unique. Otherwise, the solution is not unique, but the solution given by equation (17) has the property of having the smallest Euclidian magnitude of all the solutions. Also, in the case that the columns of **F** are linearly independent, $\mathbf{F}^T\mathbf{F} = \mathbf{D}$ is nonsingular and $\mathbf{F}^I = (\mathbf{F}^T\mathbf{F})^{-1}\,\mathbf{F}^T$. It is now seen that the iteration equation for Newton's method given in equation (13) is a special case of the solution of the general linear least-squares problem given in equation (17).

There are some implications for the numerical solution of the nonlinear least-squares problem in this observation. In the solution of the linear least squares problem, it is common to avoid direct inversion of **D**, even when it is nonsingular. If it is close to singular, there can be large numerical errors made in direct inversion. Methods avoiding direct inversion include Gram-Schmidt orthogonalization, orthogonal triangularization, and singular value decomposition. Rewriting the Newton's iteration in equation (13) in terms of the pseudoinverse in equation (17) allows one to take advantage of such procedures.

It is convenient now to discuss the relationship between Newton's method and the method of steepest descents. The method of steepest descents may be most effective when the current parameter value is far from the minimum. In this method the change in parameters is made in the direction of the negative gradient of the chi-square function (see equation 12). Symbolically this can be written

$$\delta a = \lambda F^T \cdot \delta y \tag{18}$$

where λ is a positive, scalar constant. The method of steepest descents is supposed to give quick reductions in the chi-square function because it directs the parameter vector **a** in the direction of greatest decrease in the chi-square function.

However, as the parameter vector gets closer to the minimizing value, it is better to use Newton's method because it is based on a more detailed picture of the topology of the surface than just the direction of steepest descent. These ideas are the basis for the Levenberg-Marquardt algorithm (Press et al., 1986), where a direct inversion of a modified version of the matrix **D** is performed. The modification is to add a positive constant to the diagonal of the matrix **D** (see equations 11 and 13). This gives the effect of a mixed steepest descents and Newton's method, because steepest descents can be derived by making **D** a diagonal matrix. (see equations 11, 13, and 18). The larger the constant, the closer to a steepest descents algorithm the Levenberg-Marquardt algorithm becomes, because the more diagonally dominant **D** becomes. The diagonal dominance can also help stabilize the inversion algorithm. The size of the constant to be added is determined by the relative change in the chi-square function. It is possible that a small change in function is caused by proximity to a local minimum, so the algorithm should be more in the Newton's method mode when there is a small relative change.

A variant on the Levenberg-Marquardt algorithm, based on equations (17) and (18), can be proposed.

$$\delta a = \left(F^I + \lambda F^T\right) \cdot \delta y \tag{19}$$

Again, the scalar λ is set based on the the relative changes in **a** from step to step. Neither the matrix **D**, nor its variant, is ever formed and inverted. Although we have not made use of such an algorithm, we believe that it is derived naturally from the considerations presented so far and may find future use. The program PARFIT uses the Newton's method based on equation (17), so that the inversion explicit in equation (13) is not performed.

PARFIT: GENERAL OBSERVATIONS

A few of the particulars from experience in the fitting of damped sinusoids to simulated movement data using the program, PARFIT, are discussed in this section. The computational heart of PARFIT is given by equation (17).

We found that fitting more than three parameters of a damped sinusoid to position data was not reliable. There appears to be a trade-off phenomenon between the parameters: A small change in the data would change two or three of the parameters substantially if more than three parameters were fit. (In fact, this phenomenon can be seen to occur when trying to fit three parameters. The amplification of error caused by numerical instability is discussed below in more detail.)

To eliminate the possibility that the algorithm we used to find the pseudoinverse was the source of instability, we compared a couple of algorithms. The simpler algorithm we tested was a Gram-Schmidt orthogonalization written by Rust, Burrus, and Schneeberger (1966). The more sophisticated singular value decomposition, SVD, was also tested. It was found that with five parameters the problem is inherently unstable, because the SVD gave virtually the same answers as the simpler Gram-Schmidt algorithm. Another piece of evidence that any error will be amplified despite the numerical algorithm is that the condition number given by the ratio of maximum to minimum singular values of **F** is very large when four or five parameters are being fit. The large condition number indicates that the columns of the **F** form a nearly linearly dependent set, i.e., the effects on error are nearly the same for several different parameters changes, so that there is no unique best change in parameters within the accuracy of the data and computations.

Another problem was that the ultimate best fit depends on the initial choice of parameters used to start the iteration. That is, there is no guarantee that a global minimum is reached after running the algorithm from a single initial choice for the parameters. Further, as with any iteration based on Newton's method, it is possible that there is divergence and overflow, that there is an oscillation, or that it gets stuck on a particular parameter vector and cannot proceed.

All these observations helped shape the parameter extraction program, PARFIT. One of the most important parts of this development was the decision to reduce the number of parameters from five to, at most, three. There are a variety of ways to reduce the number of parameters. We have chosen to use either the initial conditions of position and velocity (initial condition option), or initial position and final position (boundary condition option) as constraints. With these constraints the coefficient of the cosine, A, and the coefficient of the sine, B, in equation (1) are considered to be functions of the frequency, growth, constant level, and either the initial conditions or the boundary conditions.

When the initial condition option is used the cosine and sine coefficients are eliminated as variable parameters according to the initial position and initial velocity values of the data. (Set the initial position equal to $y(t, a)$ in equation (1) evaluated at $t = 0$, and the initial velocity equal to the time derivative of $y(t, a)$ in equation (1) evaluated at $t = 0$. Then solve for A and B of equation (1) in terms of damping coefficient, circular frequency, D. C. level, initial position and initial velocity (see Braun, 1983).) When it is possible, the initial velocity is determined with a centered difference, but when it is not a less desirable forward difference is used. When the boundary condition option is used, the initial position and the final position are used to eliminate the coefficients of the sine and cosine as variable parameters. (Set the initial position equal to $y(t, a)$ in equation (1) evaluated at $t = 0$, and the final position equal to $y(t, a)$ in equation (1) evaluated at t = final time. Then solve for A and B of equation (1) in terms of damping coefficient, circular frequency, D. C. level, initial position and final position.)

The use of equation (17) requires the evaluation of partial derivatives of the model function with respect to the parameters (α, ω, C). These could be estimated by finite differences, but we have chosen to evaluate them "exactly." That is, the partial derivatives are written as closed-form functions, and then these functions are evaluated as the computation requires at various values of the parameters. The form of the partial derivatives will depend on whether the initial condition option is used, or the boundary condition option is used.

Although both methods are used, there are some reasons to prefer the boundary condition option. Using the boundary condition option means that the algorithm is not relying as heavily on the initial point as in the initial condition option. Also, a finite difference does not have to be calculated in the boundary condition option and finite differences are noisier than the original data. Further, it is found that a fit with the initial conditions option would sometimes fit well in the least-squares sense, but get the wrong curvature at the end of the data record. The boundary condition option seems to avoid this problem. However, the boundary condition option is a little more temperamental, and will not always fit curves the initial condition option will fit, especially at higher damping ratios. (Damping ratio is the ratio of damping factor, α, to circular frequency, ω, in equation (1).)

It has also been found useful to reduce the number of parameters from three to two, as this avoids the problem of near linear dependence in the matrix **F**. The reduction in the number of parameters is accomplished by fixing the ratio of the damping factor to the frequency, or damping ratio. A data file can be run for several different damping ratios, and the overall best fit can be chosen to give the damping, frequency, and constant level.

Because the best fit appears to depend on the initial choice of parameter, it is found that by running the search several times from several different initial parameter settings is helpful. For two parameter searches, the algorithm is run 36 time from different initial choices, and for three parameter searches, it is run from 144 initial choices. There is no certainty that the "best" fit is attained in this way, but the empirical results seem to indicate that several of the runs come up with nearly the correct answer with each simulated data file. This multiple initial choosing has led to the use of large amounts of computer time, so to cut down on the human operator keyboard time the program has the capability to be run in batch mode.

A related question is when to stop the iteration for any run from any initial choice. An error criterion needs to be set so that the program knows that the answer it has is acceptable. This criterion is based on the size of the mean square error, and should be set according to the expected measurement error. In the computational part of PARFIT, peak-to-trough data is normalized by the range of excursion (i.e., dynamic range), so that full range is equal to one unit. In PARFIT the normalization of the square differences between the data and the fit by the square of the standard deviations in the definition of the chi-square function, as shown in equation (3), was not performed, Thus, assuming identical standard deviations, σ, at each data point, the mean square error of the fit can not be expected to be below σ^2 divided by the square of the amplitude of the data, because of the amplitude normalization just described. While there usually is no estimate of the expected standard error in the data, the numerical resolution of the computer provides a lower bound on the expected standard error.

Not all runs will produce a fit with the tightest possible error criterion, especially when in a two parameter search the fixed damping ratio is not appropriate to the data. In these cases a number of things already mentioned can happen: overflow in the computation, oscillation among particular parameter values, and a condition where there is no numerically meaningful change in the parameter values from iteration to iteration. When the program is run in batch mode, runs in which overflow occurs are automatically run again with a less stringent mean square error criterion. (See elementary numerical analysis texts for the causes of a runaway iteration using Newton's method.) This assumes that the expected standard errors have been underestimated, so that a fit can be obtained with greater mean square error. Often, runs that end because of oscillation, or because there is no movement in the parameter values from iteration to iteration, nearly meet the error criterion and provide an adequate fit.

PARFIT: SENSITIVITY TO ERROR

In very general terms, we are concerned with variability in the data and how to interpret variability in our measures or parameter fits-whether it be due to noise, or due to causes of interest. Parameter sensitivity to error must be considered because error in the data and the arithmetic are inevitable, resulting from the finite accuracy of measuring instruments and the finite precision of computation. We want to be sure that the differences in parameters in the fits of different movement curves are meaningful. For example, if two fits differ the question is whether this difference can be accounted for by expected dispersion in the data or the noise added by finite-accuracy arithmetic. If large differences in parameter values can be accounted for in this way, the procedure, including measurement, model, and algorithm, is not very useful. In the best case, we want the algorithm to be sensitive to real parameter changes, but not to noise of the measurement or numerical type. (We may be willing to give up some sensitivity to "real" changes if they are less than the empirically defined criteria, in order to suppress some noise.)

The amplification of uncertainty from the data domain to the parameter domain depends on the topology of the chi-square function. If the mapping from the data to the parameter space is very gradual, minimum points may lie in shallow valleys of the chi-square surface. This means that a small change in data, or small arithmetic error, can produce a large change in parameter value. (Note that the shallowness only has to be in one direction in the parameter space for there to be a very sensitive parameter value.) In the discussion on using more than three parameters it could be said that the valley of the chi-square surface near the minimum of the chi-square surface was too shallow when more than three parameters were allowed to vary.

The relationship between the topology of the chi-square surface, the probability density function, and the sensitivity of the parameters can be seen easily with some assumptions. Using a Taylor series expansion on the chi-square function it can be shown that near the minimum (Press, et al., 1986).

$$\chi^2(a) - \chi^2(a^*) = \frac{1}{2}\,\delta a \cdot \mathbf{D} \cdot \delta a \qquad (20)$$

$(1/2\ \mathbf{D})^{-1}$ is what is known as the formal covariance matrix. Recall that the matrix $1/2\ \mathbf{D} = \mathbf{F}^T\mathbf{F}$, where **F** is the matrix used in the pseudoinverse computation (see equations 11 and 17). Equation (20) says the matrix 1/2 **D** describes the topology of the chi-square surface near the minimum, and for this reason it is designated the *curvature matrix*. (Recall that **D** was designated the Hessian matrix in equation (6).) In the special case that the measurement errors are normally distributed, and if $\mathbf{a}^n$ denotes the best fit to the nth data set in the ensemble and **a** is the vector of true parameters, then $\delta(\mathbf{a})^n = \mathbf{a}^n - \mathbf{a}$ is normally distributed (Press, et al., 1986)

$$P(\delta a)\, da_1 \ldots d\, a_N = \text{Constant} \cdot \exp\left(-\frac{1}{4}\delta a \cdot \mathbf{D} \cdot \delta\, a\right) \qquad (21)$$

where P is the probability density function in the space of parameters.

The last two equations exhibit the connection between the geometry of the chi-square surface near the minimum and the squared standard errors under the normal distribution assumption. Because **D** is symmetric, it has an equivalent diagonal form in a rotated parameter coordinate system. If there are diagonal elements of small magnitude, relative to the largest magnitude diagonal element in the transformed **D**, then the probability density function, P(δa), will be broad in the directions (or parameters) corresponding to these small diagonal elements. These parameters will be sensitive to small changes in the data and errors in the arithmetic. Also, these directions are the directions of small curvature, or relative flatness in the chi-square surface. Large ratios in the magnitudes of diagonal elements of the transformed D are related to the degree of linear dependence in the columns of F because diagonal elements of the transformed D are the singular values of F: the larger the ratios, the closer the columns of F are to linear dependence.

TESTING THE PROGRAM

The reliability of PARFIT was tested by running it on files of simulated data for which the parameter values are known. The simulated data files were damped sinusoidal curves, with damping ratios from zero to one in steps of one-tenth. There were files of two natural frequencies, four and twelve Hz, and two initial amplitudes, 1000 units and 150 units, for each damping ratio. These files were sampled at 200 Hz. Each file was divided into "windows," which extend from the midpoint of a peak (or valley) in the data curve, to the midpoint of the next valley (or peak). Each window was fit separately, using different techniques described below. A window may contain as many as 106 data points and as few as 8 data points, and the dynamic range (i.e., range of excursion, total displacement) of data in a window could be as large as 2000 units and as little as 4 or 5 units. Because the data is damped sinusoidal, and the windows are taken from peak-to-valley, the dynamic range and number of data points depend on the damping, frequency, initial displacement, and how far along in the oscillation the window happens to be. For instance, data with smaller frequencies will have more data points in each window, as will data with a higher damping ratio. Also, the dynamic range will decrease in a given simulation as the windows are taken further into the file, because of the damping. This final trend is most pronounced in highly damped simulations. In these tests, PARFIT attempted to fit the data with an error criterion of 10^{-4} in mean square error. Often the actual error of a fit was less than this (as low as 10^{-7} for undamped data fit with an undamped curve). When this error criterion could not be met, it was raised through a series of possible values to as high as .1.

Two and three parameter fits

As a way of testing the sensitivity of parameter fits, PARFIT was run interactively to find the shape of the chi-square surface near the minimum in order to see what confidence we could have in the fits. We fit one file with damping ratio of .4 and another with damping ratio of .8. Both files had an approximate initial amplitude of 1000 units, and both had a 4 Hz natural frequency. The .4 damping ratio file contained 27 data points, and the .8 damping ratio file contained 42 data points.

We attempted to fit the simulated data with a damping ratio of .4 using three variable parameters: damping factor, frequency, and constant level. We also fit this file using fixed damping ratios of .2, .3, .4, .5, and .6, letting only frequency and constant level vary. Similarly, for the .8 damping ratio simulated data, we tried to fit with three variable parameters, as well as with two variable parameters with fixed damping ratios of .6, .7, .8, and .9. Both the initial condition option and the boundary condition option were used in all these cases.

The three parameter fits for both the initial and boundary condition options are interesting because the fits give nearly the correct answer, but we cannot be confident in the extracted parameters, at least according to the covariance matrix. In the initial condition option fit of the .4 damping ratio file, the square root of the diagonal member of the covariance matrix corresponding to the frequency parameter is almost as large as the frequency itself: the valley is relatively flat in this direction (see Figure 2). The situation is a little better for the boundary condition option, but the square root of the diagonal element of the covariance matrix corresponding to the constant level parameter is intolerably large (see Figure 3). Also, there is a large covariance between the constant level and damping factor, indicating that small changes in data would allow the parameters to trade against one another. This has been our experience. For the .8 damping ratio files the variance in the frequency, as measured by the corresponding diagonal element of the covariance matrix, is relatively high, while things seem to improve for the variance in the constant level.

The situation improves for two parameter searches with fixed damping ratios. For the .4 damping ratio file, fit with the boundary condition option and at a fixed damping ratio of .4, the square root of the diagonal element of the covariance matrix corresponding to frequency is about 20% of the best-fit frequency, and the corresponding measure in the constant level is about 10% of the dynamic range (see Figure 4). For the initial condition option the situation is not as good (see Figure 5). The square root of the diagonal element corresponding to frequency is about the magnitude of the best-fit frequency, and the corresponding measure of constant level is about one-half the dynamic range of the data.

$$\begin{array}{ccc} & & \begin{array}{ccc}\text{damping} & \text{frequency} & \text{constant level}\end{array} \\ \begin{array}{c}.36\\ .08\\ .71\end{array} & \begin{array}{c}\xleftarrow{\sqrt{\ }}\\ \longleftarrow\\ \longleftarrow\end{array} & \left[\begin{array}{ccc}\boxed{.13} & -.03 & .14\\ -.03 & \boxed{.007} & -.006\\ .14 & -.006 & \boxed{.51}\end{array}\right]\end{array}$$

	damping	frequency	constant level
fit value	.046	.114	-.013
correct value	.046	.115	0.0

(Note that the full range of data = 1.0)

Figure 2. **Covariance matrix for three parameter fit with the initial condition option on simulated data with .4 damping ratio.**

$$\begin{array}{ccc} & & \begin{array}{ccc}\text{damping} & \text{frequency} & \text{constant level}\end{array} \\ \begin{array}{c}1.03\\ .0053\\ 5.5\end{array} & \begin{array}{c}\xleftarrow{\sqrt{\ }}\\ \longleftarrow\\ \longleftarrow\end{array} & \left[\begin{array}{ccc}\boxed{1.07} & .0018 & 5.7\\ .0018 & \boxed{.000028} & .0096\\ 5.7 & .0096 & \boxed{30.1}\end{array}\right]\end{array}$$

	damping	frequency	constant level
fit value	.048	.119	.09
correct value	.046	.115	0.0

(Note that the full range of data = 1.0)

Figure 3. **Covariance matrix for three parameter fit with the boundary condition option on simulated data with .4 damping ratio.**

$$\begin{array}{ccc} & & \begin{array}{cc}\text{frequency} & \text{constant level}\end{array} \\ \begin{array}{c}.022\\ .12\end{array} & \begin{array}{c}\xleftarrow{\sqrt{\ }}\\ \longleftarrow\end{array} & \left[\begin{array}{cc}\boxed{.0005} & -.0009\\ -.0009 & \boxed{.0145}\end{array}\right]\end{array}$$

	frequency	constant level
fit value	.116	-.0013
correct value	.115	0.0

(Note that the full range of data = 1.0)

Figure 4. **Covariance matrix for two parameter fit with the boundary condition option on simulated data with .4 damping ratio.**

$$\begin{array}{ccc} & & \begin{array}{cc}\text{frequency} & \text{constant level}\end{array} \\ \begin{array}{c}.084\\ .46\end{array} & \begin{array}{c}\xleftarrow{\sqrt{\ }}\\ \longleftarrow\end{array} & \left[\begin{array}{cc}\boxed{.007} & .033\\ .033 & \boxed{.21}\end{array}\right]\end{array}$$

	frequency	constant level
fit value	.115	-.00002
correct value	.115	0.0

(Note that the full range of data = 1.0)

Figure 5. **Covariance matrix for two parameter fit with the initial condition option on simulated data with .4 damping ratio.**

For the data files with .8 damping ratio, fits using the boundary condition option also have less variance than fits using the initial condition option. The square root of the diagonal element of the covariance matrix corresponding to the frequency parameter was about three times as high for the initial condition option compared to the boundary condition option. The square root of the diagonal element corresponding to the constant level parameter was about ten times larger than that for the boundary condition option. However, the boundary condition option was highly unreliable at damping ratios of .9 and 1.0. For instance, using a fit with a fixed damping ratio of .9 on a data file whose actual damping ratio is .8 might result in catastrophic failure.

Sensitivity to noise

The simulated data files were perfectly smooth and noise-free. To get some idea of how sensitive the fitting procedure is to noise in the data, additional data files were created by adding white noise to selected files of simulated data. The files used had a natural frequency of 4 Hz and a dynamic range of 1000 units, and damping ratios of .4 or .8. Ten new files were created from each of these files by adding different noise files with maximum amplitudes of 100 units. Fixed damping ratios in .1 increments around the correct value were used to fit the noisy data files. For each damping ratio the number of simulated files that were fit best with that damping ratio are tabulated (see Table 1).

These results suggest that the boundary condition option is less sensitive to noise than the initial condition option, because the fit with the least error had a damping ratio close to the actual damping ratio of the data more often when using the boundary condition option than when using the initial condition option. The damping ratio for the fit with the least error for the boundary condition option was always within .2 of the correct damping ratio; with the initial condition option it was this close only in 16 out of 20 cases. These results are what would be expected based on the magnitudes of entries in the covariance matrices.

Accuracy of the fits

Since the two-parameter searches gave more reliable results than the three-parameter searches, we were interested in how accurately two parameter searches could be used to determine the correct damping ratio, as well as values for observed frequency with noise-free data. The method used was to fit all the simulated data files (low and high amplitude and frequency) with all damping ratios between 0 and 1 in .1 increments, and then select the fit that had the least error for each file. Each of the 44 data files was fit using the initial and boundary condition options. The damping ratio that gave the fit with the least error was selected for each combination of data file and condition option.

Table 1.

Entries give the number of noisy simulated files with damping ratio .4 that were best fit at the damping ratios listed in the top row.

damping ratios	0.	.1	.2	.3	√ .4	.5	.6	.7
# of best fits in initial condition option	2	3	1	2	1	0	1	0
# of best fits in boundary condition option	0	0	3	3	3	1	0	0

Entries give the number of noisy simulated files with damping ratio .8 that were best fit at the damping ratios listed in the top row.

damping ratios	.5	.6	.7	√ .8	.9	1.
# of best fits in initial condition option	2	2	3	3	0	0
# of best fits in boundary condition option	0	1	2	2	5	0

Because these fits were done from peak to valley, starting with the first point in the simulated data file, the first data point in the file was the first point in the window. This meant that the program could not calculate the initial velocity for the initial condition option using centered differences, so the exact value, 0., was supplied. The result was that the fits using the initial condition option were more accurate than could normally be expected. In every case the fit with the least error occurred with the correct damping ratio for the data. For the boundary condition option, the program is able to obtain the boundary conditions directly from the data. The results obtained with this option reflect the actual behavior of PARFIT with only the proviso that the data were, of course, noiseless and perfectly smooth. With the boundary condition option, the correct damping ratio had the least error in 23 (of 44) cases. In 19 other cases the fit with the least error was within .1 of the correct damping ratio, and in the remaining 2 cases the least error was found at a damping ratio .2 below the correct value.

To compare the accuracy of different fits, we looked at the accuracy of the values that PARFIT provided for natural frequency. Using the fits with the least error for each data file, and the initial condition option, the natural frequency values were never more than 0.5% away from the correct value. However, the fits using the boundary condition option sometimes gave values as much as 81% off. These highly inaccurate values occurred when fitting data whose damping ratio was .9 or 1.0. Excluding these data files, the least accurate value had an error of 11%. The initial condition option clearly gives more accurate natural frequency values for data at .9 and 1.0 damping ratios. Data files with .8 damping ratio had fits with the least error at a .8 damping ratio more often with the boundary condition than with the initial condition option. The best results overall are achieved by using the boundary condition option when the fixed damping ratio is from .0 to 0.8, and the initial condition option when it is fixed at 0.9 or 1.0. Using this combination of fits and damping ratios, the fit having the least error for each data file will always have a damping ratio no more than .1 away from the true value for the data. These fits provided natural frequency values with a mean error of less than 1%. The constant level was not examined with regard to accuracy of fit.

CONCLUSION

Using the least-squares criterion and Newton's method as programmed in PARFIT, it appears possible to fit data from peak to valley with damped sinusoidal functions. The observations above show that some information about the program's behavior is needed to select the most accurate fits. The number of parameters should be reduced from five to two by using the initial or boundary condition options and fixing the damping ratio. The data should be fit at several different fixed damping ratios using boundary conditions, except for fits with damping ratio greater than .8, when the initial condition option should be used. Then parameter values, including damping ratio, should be taken from the parameter set of the fit with the least error. The experimenter can have some confidence that the extracted parameters obtained in this way are robust in the sense that small perturbations to the data will not cause large changes in the parameter values.

REFERENCES

Braun, M. (1983). *Differential equations and their applications*. New York: Springer-Verlag.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing*. Cambridge, England: Cambridge University Press.

Rust, B., Burrus, W. R., & Schneeberger, C. (1966). A simple algorithm for computing the generalized inverse of a matrix. *Communications of the ACM, 9*, 381-387.

Stewart, G. W. (1973). *Introduction to Matrix Computation*. New York: Academic Press.

FOOTNOTES

[†] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

[1] Newton's method in one dimension amounts to finding a local linear approximation to a function, f(x), at the current parameter value, x. The root of that approximation, x*, is found, and this gives the the new estimate for the root of the function. The formula for the root of the linear approximation is given by

$$x^* = x - {}^{f(x)}/_{f'(x)}.$$

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 231-242

"Clock" and "Motor" Components in Absolute Coordination of Rhythmic Movements*

M. T. Turvey,† R. C. Schmidt,† and Lawrence D. Rosenblum††

Human subjects swung, through motions at the two wrists, hand-held pendulums of variable mass and length. Within a pair, the two pendulums could be of the same or different magnitude. The subjects were required to produce a coordinated state in which the two rhythmic units oscillated at a single common period. Fifty four conditions of absolute coordination, that is, 54 different pairs of wrist-pendulum systems, were investigated for each of three subjects in the course of six sessions. Each condition of absolute coordination was conducted in the out of phase mode and at the single most comfortable period. The period variances of the right and left systems in the 162 instances of absolute coordination were analyzed according to a method that assumes that a timekeeper function and a motor implementation function contribute independently to the variance in the periodic timing of a rhythmic movement. The major findings were that in absolute coordination: (a) a system's "motor" variance, but not its "clock" variance, depended on the deviation of the period of absolute coordination from the system's characteristic period; (b) right and left "clock" variances were related; and (c) neither the "motor" variances nor the "clock" variances were affected by deviations in the phase relation from 180 degrees. The results were discussed in terms of their implications for interpreting von Holst's notions of maintenance tendency and magnet effect and, more generally, the neural and dynamical basis of absolute coordination.

It is often proposed that two functions underlie the rhythmic limb movements typifying locomotion and that these two functions are subserved by different neural elements (von Holst, 1937, 1939/1973; Pearson, 1976). A timekeeping function is said to be provided by single central cells or populations of central cells that produce a periodic signal; a motor function is said to be provided by populations of central cells which, under this periodic control, transmit impulses to the muscles. A way of conceptualizing this neural division of labor in terms of the variance of a rhythmic movement's period has been developed by Wing and Kristofferson (1973). Following a suggestion by Ten Hoopen and Reuver (1967), Wing and Kristofferson adopt the strategy of accounting for the sequential dependence of the durations of movement cycles treated as a time series. At a regular interval C the timekeeper is assumed to emit a signal that initiates a response in the motor components. A short delay intervenes between the initiation of this response and the overt, measured behavior. Each period τ_j of rhythmic movement can be viewed, therefore, as the sum of the timekeeper's cycle time (C_j) plus the difference in the motor delays (MD_j, MD_{j-1}) connected with the responses that initiate and terminate the period:

$$\tau_j = C_j + MD_j - MD_{j-1}. \quad (1)$$

If the timekeeper intervals and the motor delays are independent random variables with variances of s_c^2 and s_{md}^2, respectively, then it follows that the period variance s_τ^2 is given by:

$$s_\tau^2 = s_c^2 + 2s_{md}^2. \quad (2)$$

This research was supported by a grant from the Office of Naval Research (Contract N00014-83-C-0083) awarded to the Haskins Laboratories, and by a grant from NSF (BNS-8811510) awarded to the first author. The writing of the manuscript was supported, in part, by a James McKeen Cattell Fellowship awarded to the first author.

The period variance $s\tau^2$ can be obtained directly from the rhythmic movement data. The component variances $s_c{}^2$ and $s_{md}{}^2$ can be estimated from the covariance function of the successive cycle periods. The central predictions of Wing and Kristofferson (1973) are that (a) temporally abutting cycles will be negatively correlated and (b) temporally nonabutting cycles, those separated by one cycle or more, will be uncorrelated. The negative sign of the correlation between adjacent cycles means that if one cycle period is by chance shorter than average the next cycle period will be longer than average, and vice versa (Wing, 1980). The negative dependence does not point to a fed-back influence (in the form of temporal information) from one cycle to the next. Rather, it is seen more simply as an inevitable result of the inherent delays in the motor system: A large motor delay in the j-1 cycle will induce a long j-1 cycle period and a short j cycle period. The lag 1 covariance is proposed, therefore, as a measure of the motor delay variance :

$$\mathrm{cov}(\tau_j\ \tau_{j\text{-}1}) = -\ s_{md}{}^2 \tag{3}$$

where

$$\mathrm{cov}(\tau_j \tau_{j\text{-}1}) = \Sigma\ (\tau_j - \tau)(\tau_{j\text{-}1} - \tau)/(N\ \text{-}1) \tag{4}$$

and N is the number of cycles. Given that the lag 0 covariance is a measure of $s\tau^2$ (τ being the mean value of the periodic time), that is,

$$\mathrm{cov}(\tau_j \tau_j) = \Sigma(\tau_j - \tau)^2/N, \tag{5}$$

then it follows that

$$\mathrm{cov}(\tau_j\ \tau_j) = s_\tau{}^2 = s_c{}^2 + 2s_{md}{}^2. \tag{6}$$

The lag 1 autocorrelation between adjacent cycle times is expressed most usefully by the ratio of (3) to (6):

$$\mathrm{cor}(\tau_j \tau_{j\text{-}1}) = \text{-}1/[2 + s_c{}^2/s_{md}{}^2]. \tag{7}$$

On the average, the predicted negative lag 1 autocorrelation between adjacent cycles is expected to be between 0 (timekeeper variance is very large relative to motor variance) and -0.5 (timekeeper has no variability). Combining the equations for the lag 1 autocorrelation and the period variance $s_\tau{}^2$ yields measures of the "clock" and "motor" variances, namely,

$$s_c{}^2 = s_\tau^2 + 2[s_\tau^2 \times \mathrm{cor}(\tau_j \tau_{j\text{-}1})] \tag{8}$$

and

$$s_{md}{}^2 = \text{-}s_\tau^2 \times \mathrm{cor}(\tau_j \tau_{j\text{-}1}), \tag{9}$$

respectively.

In this article we use the analysis of Wing and Kristofferson to evaluate the behavior of clock and motor components in absolutely coordinated rhythmic movements. In absolute coordination two or more limbs oscillate at the same frequency. von Holst (1937, 1939/1973) interpreted absolute coordination at a freely elected frequency to be the result of two inclinations working in direct opposition, namely, the inclination of an individual rhythmic unit to move at its own characteristic frequency (the maintenance tendency), and the inclination of an individual rhythmic unit to move at the characteristic frequency of another rhythmic unit (the magnet effect). If the maintenance tendency dominated then the units (limbs or fins) would oscillate independently of one another at different frequencies. If the magnet effect dominated then the units would oscillate at one and the same frequency. With respect to the latter condition, however, von Holst (1937, 1939/1973) concluded that "...the 'struggle' between the magnet effect and the maintenance tendency continues *in latent form* even after attainment of absolute (successful) coordination" (p. 65).

A methodology introduced by Kugler and Turvey (1987) permits wide systematic control over both the characteristic frequencies of individual rhythmic units and the departures from characteristic frequencies within the context of stable, absolute coordinations. The individual rhythmic movement unit is defined by a hand-held pendulum together with the neuromuscular processes by which the pendulum is moved rhythmically about an axis in the wrist at a comfortable tempo. The characteristic frequency of such a rhythmic unit—a wrist-pendulum system—will be determined in largest part by the pendulum's magnitude. Consequently, the systematic variation of pendulum size will result in the systematic variation of characteristic frequency (Kugler & Turvey, 1987; Turvey, Schmidt, Rosenblum, & Kugler, 1988). Absolute coordination is achieved when two wrist-pendulum systems (left and right) move rhythmically together under the requirements (instructions to the subject) that the two systems (a) maintain a particular phase relation and (b) move at the single common period that is most comfortable. For two wrist-pendulum systems of different characteristic periods, meeting requirement (b) means that one of the two systems moves, or both of the systems move, at a nonpreferred period. To manipulate the difference between the two absolutely coordinated wrist-pendulum systems—specifically, to manipulate the size difference—is to control the degree to which the period of an individual wrist-pendulum system will depart from its characteristic period.

The absolute coordination of two wrist-pendulum systems forms a coordinative structure analyzable into three levels (Kugler & Turvey, 1987; Turvey, Rosenblum, Schmidt, & Kugler, 1986). The upper level comprises the constraints the subject brings to bear in accordance with the experimental instructions. The lower level comprises the two individual wrist-pendulum systems. The middle level comprises a cooperative state in which the two systems act together—namely, the state of absolute coordination. The lower level units, in order to satisfy cooperativity, deviate from the rhythmic parameters they would exhibit in isolation; to achieve comfortable absolute coordination with its partner, an individual wrist-pendulum system must oscillate at a frequency that is faster or slower than preferred. One consequence of this deviation ought to be increased variability in the parameters of oscillatory behavior (Kugler & Turvey, 1987). It has been observed that when a wrist-pendulum system in absolute coordination oscillates at frequencies other than its characteristic frequency, the fluctuations in the system's frequency and amplitude are magnified (Rosenblum & Turvey, 1988).

Returning to the language of clocks and motors, two hypotheses can be advanced about the state of absolute coordination. The first hypothesis is that the lower level comprises two distinct clock/motor complexes with the middle level arising from the mutual adjustment of these two complexes in the sense of von Holst's magnet effect. According to the first hypothesis, both the motor and clock variances of the two systems would be expected to reflect the maintenance tendency, with both kinds of variances increasing as a function of the degree to which the period of absolute coordination differs from the systems' characteristic periods. The basis of the preceding prediction is a particular interpretation of cooperativity, namely, that of component units compromising their preferred states in order to achieve a given effect. The second hypothesis is based on an interpretation of cooperativity as the formation of a unitary process. According to this second hypothesis, the fact that the lower level units oscillate at an identical, or nearly identical, frequency implies a unitary timekeeper—a single, dynamical organization underlying the equal durations of right and left cycles. That is, absolute coordination is not so much a type of coupling of two local clocking mechanisms as it is the formation of a new, more globally defined clock that provides a reference frame for the intracycle patterning of muscular activities in the lower level units (Turvey et al., 1986). According to this second hypothesis, the clock variances of right and left systems in comfortable absolute coordination should be independent of how far the systems deviate from their characteristic periods. The motor variances, in contrast, should depend on the size of the deviation on the assumption that a characteristic period identifies the preferred pace of the intracycle motor pattern.

In short, we propose and test in the present article two hypotheses about comfortable absolute coordination. The contrast between the two hypotheses may be summarized as follows: (a) The first hypothesis predicts that both clock and motor variances should increase with deviations from a wrist-pendulum system's characteristic frequency; the second hypothesis predicts that only the motor variance should increase; (b) The first hypothesis predicts that the clock variances of right and left wrist-pendulum systems should be uncorrelated; the second hypothesis predicts that the clock variances of right and left wrist-pendulum systems should be correlated.

Evaluating these two hypotheses has significance for recent theorizing about the neuronal networks underlying locomotion. It is usually contended that a single circuit is responsible for both the patterning of motor activity within a limb or fin cycle and the generating of the cycle-to-cycle timing of the locomotory rhythm (Grillner, 1975). In contrast, Lennard (1985) has suggested that there is a functional separation of the networks subserving intracycle pattern generation and the timing of the locomotory activity. The oft-cited central pattern generator would be limited to producing the proper phase relations between muscles within each cycle; an independent central timing network would be responsible for initiating each cycle and, therefore, for organizing the locomotor rhythm. Lennard (1985) hypothesizes that the intracyle pattern generator may be conceptualized as an oscillatory network that is entrained by the timekeeping network, or as a heavily damped oscillator (one-cycle pattern generator) driven on a cycle-to-cycle basis by the timekeeping network. We suggest that support for Lennard's (1985) functional separation of timing and pattern generating networks would be provided by a clock/motor analysis favoring the second of the two hypotheses advanced above.

Method

The foregoing predictions are evaluated with the data from a study consisting of six separate sessions (or experiments) conducted over a 21 month interval with the same three subjects. The study was designed to address simultaneously a number of different issues concerning absolute coordination. Some methodological details and results of the study have already been reported (Rosenblum & Turvey, 1988; Turvey et al., 1988). The emphasis in what follows is on the features of the experimental design that bear directly on questions concerning the timekeeping and motor components of rhythmic systems in absolute coordination, for example: How are the clock variances of the two systems related? How are the motor variances of the two systems related? How are the clock and motor variances of a system affected by deviations from the system's characteristic period?

Subjects

The three subjects were 21, 24 and 41 years of age at the start of the study. Each subject was right handed. The subjects weighed as follows: Subject 1, 73.5kg, Subject 2, 74.9kg, Subject 3, 78.0kg.

Materials

The pendulum construction, recording procedure and method for computing the equivalent simple pendulum lengths were as described in Kugler and Turvey (1987), Turvey et al. (1986), and Rosenblum and Turvey (1988).

In each session of the study the right wrist pendulum was of fixed magnitude. Different conditions of absolute coordination were produced by varying the magnitude of the left wrist pendulum. The characteristic frequencies of the (constant) right systems considered as strictly gravitational pendulums, and the range of characteristic frequencies for the (variable) left systems considered as strictly gravitational pendulums, are reported in Table 1 for each subject in each of the six sessions of the study. Because the masses of the subject's hands enter into the computations of these simple gravitational pendulum equivalents of wrist-pendulum systems (Kugler & Turvey, 1987), the characteristic frequencies differ for the three subjects.

Procedure

Subjects (run one at a time) sat on a stool with their feet planted firmly on a foot stand. They were given extensive verbal instructions in the first session. Each subject was instructed to hold his forearms parallel to the ground-plane and to gaze straight-ahead without looking at either wrist-pendulum system. He was asked to oscillate the pendulums forward and backward smoothly using only the wrist joints while gripping each pendulum's handle so as to have complete control over the entire swing. The subject was told to swing the pendulums in alternate fashion (one pendulum moving forward simultaneously with the other pendulum moving backward) with a single, most comfortable common period. Importantly, before the recording of each trial, the subject was given as long as needed to settle on a tempo that he felt was comfortable and stable. Each recorded trial lasted for 15 s, except in Session 6 in which the trials were of 10 s duration. The set of instructions was repeated to subjects in briefer form for experimental sessions S-6 and subject behavior was monitored closely throughout the six sessions.

In Sessions 2, 3, 4 and 5 there were eight absolute coordination conditions and one single right wrist-pendulum condition. The single right wrist-pendulum condition in each session permitted the determination of τ_0, the right system's characteristic period. To reiterate, the right system was held constant across the absolute coordination conditions of each session.

Table 1. *Characteristic frequencies (in Hz) of the constant right systems, and the range of the characteristic frequencies of the variable left systems, when the systems are considered as strictly gravitational pendulums.*

	Subject 1		Subject 2		Subject 3	
	right	left range	right	left range	right	left range
Session 1	0.89	1.09 - 0.74	0.89	1.17 - 0.74	0.90	1.19 - 0.75
Session 2	0.89	1.45 - 0.62	0.89	1.47- 0.62	0.90	1.52 - 0.63
Session 3	1.08	1.45 - 0.62	1.09	1.47 - 0.62	1.11	1.52 - 0.63
Session 4	0.76	1.45 - 0.64	0.76	1.47 - 0.64	0.77	1.52 - 0.64
Session 5	1.27	0.97 - 0.57	1.29	0.98 - 0.57	1.32	0.99 - 0.57
Session 6	1.43	0.97 - 0.57	1.45	0.98 - 0.57	1.50	0.99 - 0.57

In Sessions 1 and 6 there were eleven absolute coordination conditions and one single right wrist-pendulum condition. In Sessions 1, 2 and 5 there were eight trials per condition; in Sessions 3 and 4 there were four trials per condition; and in Session 6 there were six trials per condition. For each session, the total trials were divided into a number of blocks each of which involved one trial of every condition type. Trials within each block were given a random ordering and this ordering was different for each subject. The sessions varied in length from about 1 1/2 to 3 1/2 hours (depending on the number of trials involved) including a fifteen minute break that occurred halfway through each session. The six sessions spanned 21 months with an average interval of three months between sessions.

Analyses

The digitized trials were analyzed by the AngS and Vongraph waveform analysis programs at Haskins Laboratories. The peaks and valleys were "picked" using the criteria of the lowest and highest points of each cycle. Peak to peak durations were used in the calculations pertaining to periodic timing within a trial. The difference of one wrist-pendulum system's peak with respect to the other wrist-pendulum system's peak was used to calculate the mean phase relation within a trial (Yamanishi, Kawato, & Suzuki, 1979).

For each of the three subjects the mean period and period variance were computed for each trial of each condition. For each subject, the means of these trial values were used as the estimates of a condition's τ and s_τ (Chatfield, 1975). Similarly, for each of the three subjects, the mean lag 1 autocorrelation was computed for each trial. For each subject, the mean of these trial values was used as the estimate of a condition's lag 1 autocorrelation.

For some of the analyses the period fluctuation measure introduced by von Holst (1937, 1939/1973) was used. This measure has the form:

$$\Delta T = \sqrt{\left(\left[\begin{array}{l}\left\{\sum (T_{pv}\text{-}MT_{pv})^2/MT_{pv}\right\} \\ +\left\{\sum (T_{vp}\text{-}MT_{vp})^2/MT_{vp}\right\}\end{array}\right]\Big/N\right)}, \qquad (10)$$

where ΔT is the relative fluctuation magnitude for a particular trial; T_{pv} is the peak-to-valley duration of a particular cycle; MT_{pv} is the mean peak-to-valley duration across all cycles of a given trial; T_{vp} is the valley-to-peak duration of a particular cycle; T_{vp} is the mean valley-to-peak duration across all cycles of a given trial; and N is the number of cycles per trial. The von Holst fluctuation measure is a relative fluctuation measure that takes into account the fact that the flexion and extension phases of a cycle are often asymmetric. As a consequence, the component durations of a movement cycle (in waveform terms, the peak-to-valley component and the valley-to-peak component) can be unequal. Also, by scaling each variance score to the mean period of the trial from which it was taken, the measure takes into account the fact that from one observation of a given steady-state rhythmic movement unit to the next observation, the mean periodicity will vary.

Results

As identified in the introduction, the central predictions of Wing and Kristofferson (1973) are that (a) temporally abutting cycles will be negatively correlated with an average lower bound of -0.5, and (b) temporally nonabutting cycles, those separated by one cycle or more, will be uncorrelated. Application of the clock and motor analysis rests on the satisfaction of these two predictions. They follow directly from the central assumption of independent clock and motor processes. To evaluate the two predictions, autocorrelation coefficients were computed for each trial on the durations of cycles at different lags, with the largest lag defined by $N/4$ where N is the number of cycles in a trial (Chatfield, 1975). (Given that trial durations were fixed, the number of cycles observed within a trial varied with the conditions of absolute coordination.) The mean autocorrelation coefficients for the left and right systems of each subject in each session and their significances are presented in Table 1 for lags 1-3. Higher lags under the $N/4$ cut-off criterion were rare and none were significant. As can be seen from inspection of Table 2, the data from the three subjects in the six sessions conform extremely well to the predictions of Wing and Kristofferson's theory. Lag 1 autocorrelations were predominantly negative, significant, and greater than -0.5; higher lag autocorrelations were predominantly insignificant.

In each experiment for each subject, the right system lag 1 autocorrelation coefficients were averaged across the trials for a given condition. Figure 1 presents the 162 lag 1 values of the right systems in absolute coordination (54 conditions × 3 subjects) as a function of τ/τ_0. This ratio of the period of absolute coordination to the right system's characteristic period defines the deviation of the right system from its preferred cyclic state. Two main features are to be noted.

Table 2. *Serial autocorrelations at lags 1-3 for the three subjects in the six sessions.*

	Left Hand			Right Hand		
	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3
Session 1						
Subject 1	-.245**	-.056		-.197**	-.029	
Subject 2	-.249**	-.100*	-.056	-.201**	-.037	-.002
Subject 3	.068*	-.022	-.073	.066	.009	-.033
Session 2						
Subject 1	-.217**	.001	.008	-.199**	.049	.002
Subject 2	-.296**	-.048	-.067	-.280**	-.009	-.009
Subject 3	-.133**	.040	.015	-.033	-.065	-.060
Session 3						
Subject 1	-.114*	.008	.041	-.024	.055	-.027
Subject 2	-.213**	-.077	-.001	-.213**	.033	-.070
Subject 3	-.114	.138*	-.018	-.103	.060	-.060
Session 4						
Subject 1	-.136*	-.077	-.116	-.089	.003	.051
Subject 2	-.378**	.060	-.125	-.314**	-.024	.104
Subject 3	-.185*	.073	-.035	-.035	.002	-.057
Session 5						
Subject 1	-.217**	-.057	-.004	-.271**	-.053	-.027
Subject 2	-.291**	-.027	.051	-.286**	.002	-.116*
Subject 3	.001	.027	-.021	-.159**	.023	.001
Session 6						
Subject 1	-.409**	-.098		-.328**	.039	
Subject 2	-.198**	-.034	-.028	-.342**	-.010	-.039
Subject 3	-.255**	-.018	.005	-.311**	-.017	-.020

*$p < .05$, **$p < .001$. Empty cells are due to insufficient cycles for computing lag 3.

First, the lag 1 autocorrelations are significantly and negatively related to the degree to which a system's period deviates from its characteristic period. The implication of this outcome is that with increasing deviation from the characteristic period, motor variance was amplified more than clock variance (see Equation 7). Second, although the lag 1 autocorrelations were overwhelmingly negative and in the range of 0 to -0.5 as predicted by the theory of Wing and Kristofferson, some were positive. With the right lag 1 autocorrelations greater than 0 and less than -0.5 removed (a total of 23 cases out of 162), the linear regression remains significant and of negative slope: $r^2(138) = 0.133$, $p < 0.0001$, intercept = -0.098, slope = -0.089.

Figure 2 presents the relation between the 162 lag 1 autocorrelations of the right systems and the corresponding 162 lag 1 autocorrelations of the left systems. Again, there are two main observations to be made. First, there was evidently a significant tendency for left lag 1 autocorrelations to be large when right lag 1 autocorrelations were large. The implication from the structure of Equation 7 is that motor and/or clock variances across the two systems were linked. Second, although the left lag1 autocorrelations were overwhelmingly negative and in excess of -0.5, there were a number of violations to the expectation from the theory of Wing and Kristofferson; 30 of the 162 autocorrelations were positive and one was less than -0.5. With left and right lag 1 autocorrelations in excess of 0 and less than -0.5 excluded, the linear relation between left and right systems remained significant and positive: $r^2(119) = 0.168$, $p < 0.0001$, intercept = -0.134, slope = 0.39.

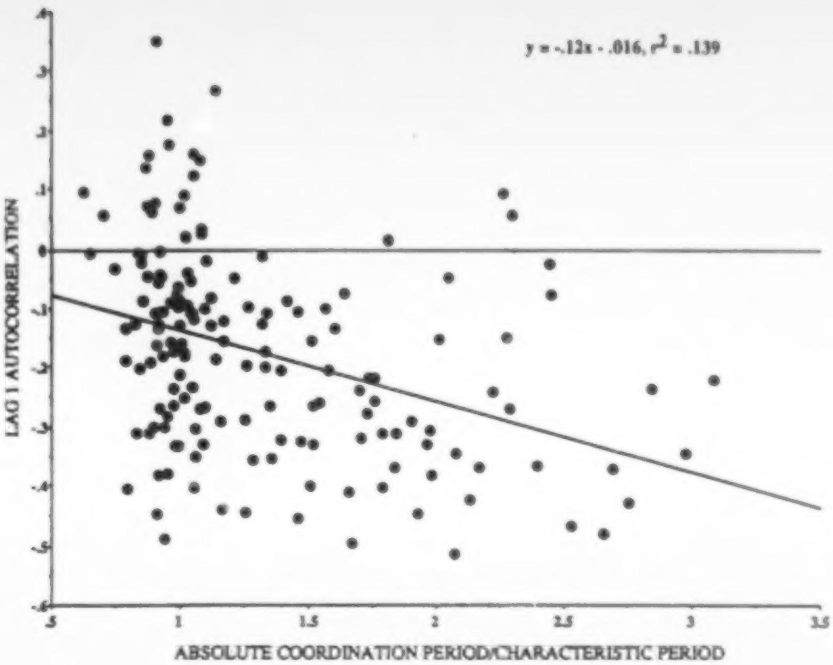


Figure 1. **Lag 1 autocorrelations of the right systems expressed as a function of deviation from characteristic period. The x-axis is the ratio of the period of absolute coordination (τ) to the period a right system oscillates at in isolation (τ_0).**

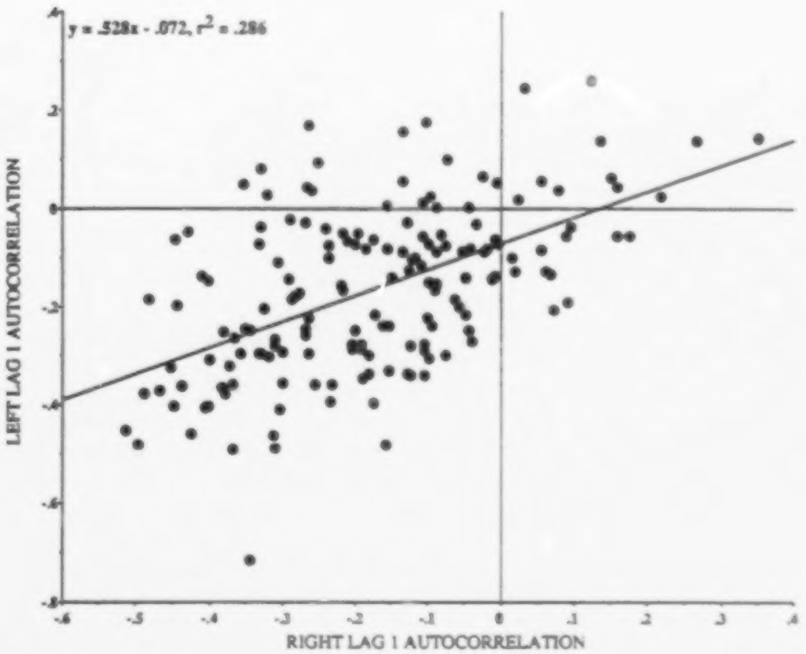


Figure 2. **Linear regression of left systems' lag 1 autocorrelations on right systems' lag 1 autocorrelations.**

For each condition for each subject, motor and clock variances (and standard deviations) were computed for both right and left systems according to Equations 8 and 9. (By definition, these computations were restricted to the range: 0 > lag 1 autocorrelation > -0.5.) Two multiple regressions, focusing on the motor and clock components of the right systems, were conducted on the overall data and repeated for the individual subject data. In each condition of absolute coordination, the two systems oscillated at a single common period and at a bounded phase relation (these features are documented in Rosenblum and Turvey (1988)). The period mean was 1.117 s; the period range was 0.604 s to 1.736 s. The phase mean was 189.4 deg; the phase range was 118 deg to 220 deg. Also, in each condition of absolute coordination the right system oscillated at a period other than its measured, characteristic period. The mean ratio of period to characteristic period was 1.32; the range of the ratio was 0.63 to 3.09. In one multiple regression, the right motor standard deviation (M_R) was regressed simultaneously on distance from characteristic period (τ/τ_0), period (τ), mean phase relation between right and left systems (ϕ), left motor standard deviation (M_L), and left clock standard deviation (C_L). In the other multiple regression, right clock standard deviation (C_R) was regressed on the same five independent variables. The coefficients for the independent variables and their associated probabilities are presented for both regressions on the overall data in Table 3. The r^2 for the motor analysis was 0.601; the r^2 for the clock analysis was 0.413. The total number of degrees of freedom was 111. The attrition from 162 was due to the combination of right and left lag 1 autocorrelations that were outside the theoretically accepted range. Conditions where either the left or right lag 1 autocorrelation was unacceptable were automatically excluded from the analysis. With respect to right motor fluctuations, τ/τ_0, τ, and M_L made significant, independent contributions in the overall data. That is, the motor component of a right system in absolute coordination was affected by the deviation of the period of absolute coordination from the system's characteristic period, by the duration of the period of absolute coordination, and by the motor component of the left system with which it was absolutely coordinated. With respect to right clock fluctuations, only C_L made a significant independent contribution. Importantly, right clock fluctuations in absolute coordination were not affected by the right system's deviation from its characteristic period.

Table 3. *Multiple regression analysis of component variances of right systems.*

	Variance Type			
Independent	MOTOR		CLOCK	
Variable	Coefficient	Probability	Coefficient	Probability
τ/τ_0	9.36	$p < .0006$	-0.09	ns
ϕ	-11.09	ns	6.87	ns
τ	33.65	$p < .0001$	14.30	ns
M_L	0.20	$p < .007$	0.01	ns
C_L	-.01	ns	0.60	$p < .0001$

ns = nonsignificant at .05 level.

On the basis of the overall analyses it would appear that in absolute coordination (a) there is a motor subcomponent common to the two oscillatory systems (M_R is linked to M_L); (b) there is a motor subcomponent specific to the right oscillatory system (M_R is affected by τ/τ_0) and, presumably, a motor component specific to the left oscillatory system; and (b) there is a single, common timekeeper (C_R is linked to C_L but is unaffected by τ/τ_0). The analyses of individual subject data are strongly supportive of (b) and (c), and weakly supportive of (a). In the M_R multiple regression, only τ/τ_0 was significant for Subjects 1 and 3; for Subject 2, τ/τ_0, τ, and M_L were significant. In the C_R multiple regression, only C_L was significant for Subjects 1 and 2; for Subject 3, only the τ beta weight reached significance in an analysis for which the r^2 was insignificant.

The periods of the left systems in isolation were not measured in the six sessions of the study. To analyze the left systems more completely, the periods of these systems considered as strictly gravitational pendulums were used as estimates of their characteristic periods (Table 1). Let a system's gravitational period be designated by τ_g. Then the ratio τ/τ_g provides a deviation measure comparable to τ/τ_0. The mean value of τ/τ_g for the left systems was 0.951 with a range of 0.781 to 1.608. Two overall multiple regressions were computed focusing on the motor M_L and clock C_L variances of the left systems, with τ/τ_g, τ, ϕ, M_R and C_R as independent variables. The results reproduced closely the results of the right systems' analyses presented in Table 3. For the M_L regression, the beta weights of τ/τ_g, τ and M_R were significant ($p < 0.02$, $p < 0.004$, and $p < 0.01$, respectively), those of ϕ, and C_R were insignificant ($p > 0.05$). For the C_L regression, only τ and C_R proved to be significant ($p < 0.02$, and $p < 0.0001$, respectively).

We present two other analyses for completeness. First, it is possible to gain a more precise impression of the rate of growth in right motor variance as a function of τ/τ_0 by expressing each right system's motor variance in absolute coordination in units of its motor variance in isolation. Figure 3 presents normalized motor variance, that is, (motor variance at τ)/(motor variance at τ_0) as a function of τ/τ_0 in double logarithmic coordinates. The resulting power function is: normalized motor variance = $1.26(\tau/\tau_0)$ (Grillner, 1975; Pearson, 1976; Rosenblum & Turvey, 1988) with 95% limits on the exponent of 3.27 (lower) and 4.51 (upper). Second, in the report of Rosenblum and Turvey (1988) the period fluctuation measure was that introduced by von Holst (1937, 1939/1973). Figure 4 shows in double logarithmic coordinates the dependencies of right motor and clock variances, as calculated with the von Holst relative fluctuation measure (Equation 10), on deviation from characteristic period in the region $\tau/\tau_0 \geq 1$. Rosenblum and Turvey (1988) observed that the growth of relative fluctuations in period was symmetrical about $\tau/\tau_0 = 1$ but that it was statistically more reliable in the $\tau/\tau_0 \geq 1$ range than in the lower range largely because of the experimental limitation on the number of $\tau/\tau_0 \leq 1$ values. The r^2 values for the motor and clock linear regressions presented in Figure 4 were 0.252 ($p < 0.0001$) and 0.046 ($p > 0.05$), respectively. That is, motor variance increased with deviations from characteristic period, but clock variance did not, confirming the main observation with motor and clock variances computed conventionally.

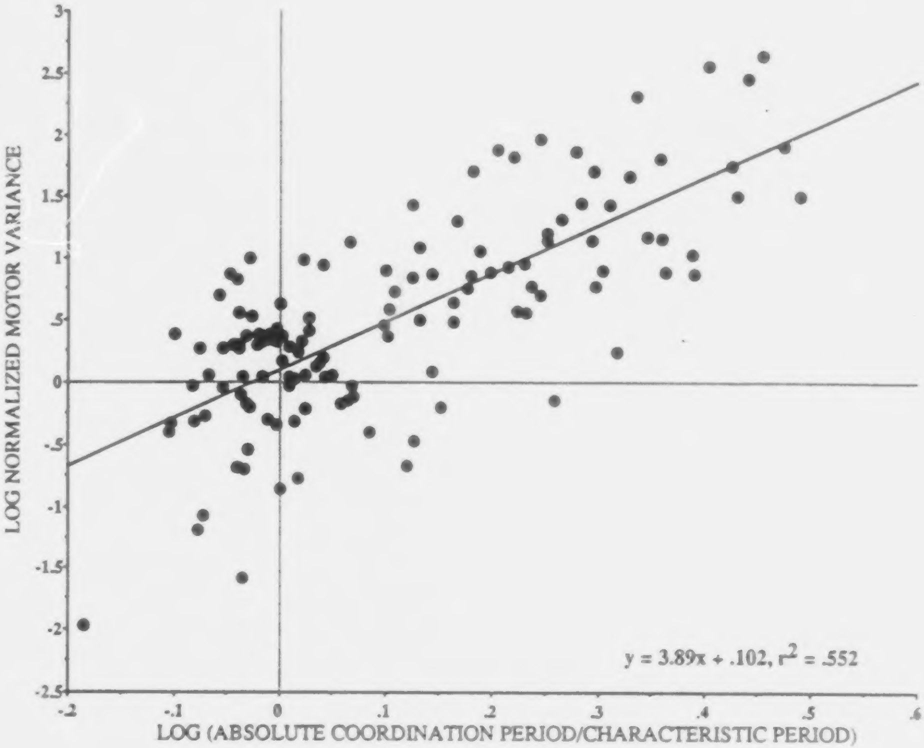


Figure 3. Plot in double logarithmic coordinates of the ratio of motor variance at period of absolute coordination/motor variance at characteristic period (y-axis) against the ratio of period of absolute coordination/characteristic period (x-axis).

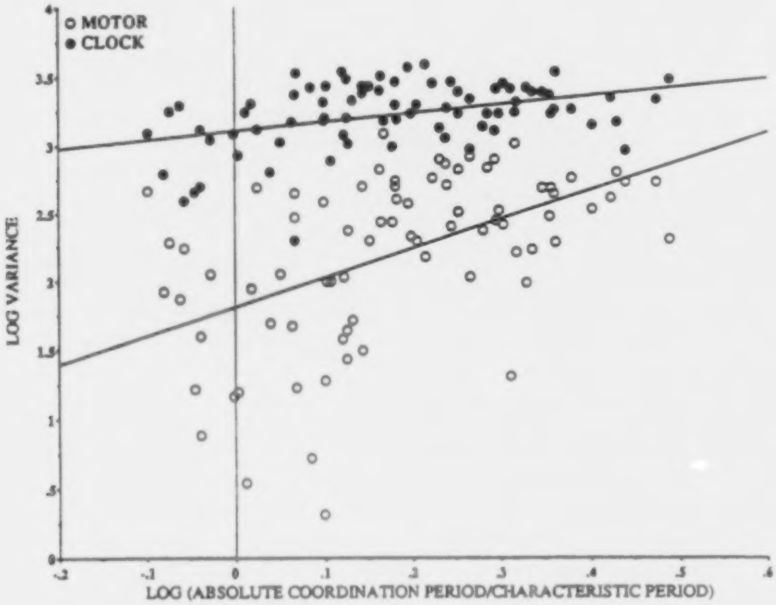


Figure 4. Plot in double logarithmic coordinates of right systems' clock and motor variances as a function of deviation from characteristic period. Clock and motor variances represented in this figure are computed from the von Holst period variance measure (see text for details).

Discussion

The present results support the second of the two hypotheses proposed in the introduction: The clock variances of right and left systems were correlated, and only the motor variance was affected by a wrist-pendulum system's departure from characteristic frequency. The implication is that absolute coordination is to be understood as the assembling of a single common clocking process and not as the linking of already established clocks. Additionally, the present results complement those of an experiment by Turvey et al. (1986) in which absolute coordination of wrist-pendulum systems moving forward simultaneously (in phase, symmetric) was compared with absolute coordination of wrist-pendulum systems moving in opposite directions simultaneously (out of phase, asymmetric). The Turvey et al., (1986) experiment revealed that the contrast of phase symmetry affected right and left clock variances (out of phase clock fluctuations were largest) but did not affect right and left motor variances. When the results of the two experiments are combined a clear cut double dissociation emerges: Clock variance is affected by phase symmetry but not by deviation from characteristic frequency; motor variance is affected by deviation from characteristic frequency but not by phase symmetry.

What might be inferred from the different dependencies of the motor and clock variances? In swinging a hand-held pendulum, radial flexion opens the wrist joint and ulnar flexion closes the wrist joint. This intracycle pattern of muscle activations can be said to comprise the motor component of the Wing and Kristofferson analysis. Wing and Kristofferson's motor component, therefore, is conceptually close to what is commonly termed a central pattern generator (CPG). Functional models of the neural networks underlying locomotion typically ascribe to a single circuit, namely, the CPG, the responsibility for both (a) producing the proper phase relations between muscles within a cycle and (b) generating the cycle-to-cycle timing (Grillner, 1975). The identification of (a) as the source of motor variance, and the different dependencies of clock variance, suggest that (a) and (b) are separate functions and only function (a) should be assigned to the CPG. This conclusion comports with the separation of timing and pattern generating functions suggested by Lennard (1985) and described in the introduction.

Positive lag 1 autocorrelations

Although the lag 1 autocorrelations in the present study were overwhelmingly negative and in the range of 0 to -0.5 as predicted by the theory of Wing and Kristofferson, some were positive.

These positive values were mainly in the vicinity of periods equal to or less than the characteristic period (see Figure 1). The analysis of Wing and Kristofferson attributes period variance wholly to neural factors. There is good reason to expect, however, that where a body segment is oscillating relatively fast, additional non-neural factors enter into the determination of period variance. One such additional non-neural factor is an elastic recoil, specifically, the storage of mechanical energy that occurs when contracted muscles are stretched forcibly immediately before shortening (Cavagna, 1977). If this factor is exploited in the performance of a rhythmic movement, and if it's magnitude is variable from cycle to cycle, then the variance in period will be due partly to the variance in elastic recoil. Adding further sources of variance to the model of Wing and Kristofferson can be expected to affect the equation set in non-trivial ways, resulting, perhaps, in predictions of positive lag 1 autocorrelations under certain conditions of oscillation.

Component variances and phase deviations from 0 and 180 degrees

In the present study, subjects executed their absolutely coordinated motions out of phase. As noted, however, they usually produced a phase relation greater or smaller than 180°. If any given movement phase relation of $x°$ reflects a neural activation phase relation of $x°$, and if 0 and 180° identify highly preferred phase relations in the sense of "attractors" (Haken, Kelso, & Bunz, 1985), then effects of deviations from 0 and 180° on the component variances should have been observed. The absence of such effects suggests that the different phase relations exhibited in the absolute coordination of wrist-pendulum systems may not be indicative of different neural processes—for example, a right-left movement phase relation that deviates from 180 degrees may not identify a right-left neural activation phase relation that deviates from 180° (Rosenblum & Turvey, 1988).

Conclusions

Absolute coordination is perhaps the most basic of all movement coordinations (it typifies walking, running, hopping, swimming, and flying), yet the principles behind it remain largely unknown. In the present article we have applied an analysis developed by Wing and Kristofferson (1973; Wing, 1980) to data on the absolute coordination by human subjects of hand-held pendulums. The analysis assumes an independence of timekeeping and motor implementation functions. The major expectations from this assumption with respect to a time series treatment of sequential cycle durations were confirmed. Moreover, the results reported in this article and elsewhere Turvey et al., (1986) identify in absolute coordination reliable and theoretically interesting differences between the two functions as they are measured by the method of Wing and Kristofferson. One major overarching conclusion, therefore, is that the independence assumption and the details of its interpretation may map onto real aspects of rhythmic movement.

Other conclusions follow from the preceding and bear on efforts to describe absolute coordination in neural and dynamical terms. First, there appear to be two isolable and independent aspects of the motor implementation component. One aspect can be linked to the intracycle phasing of muscle contractions. This aspect is sensitive to the characteristic period and becomes increasingly more variable with increasing deviations of the operating period from the characteristic period. It expresses the maintenance tendency as defined by von Holst (1937, 1939/1973). The other, perhaps less prominent aspect of motor variance, may be linked to intentional modifications in rhythm designed to maintain a particular constraint (e.g., "move the two limbs together at the most comfortable frequency"). Unlike the first motor aspect, which is specific to each of the units in absolute coordination, the second motor aspect is common to the two units. Second, it appears that absolute coordination is characterized by a single timekeeping function. This function expresses the magnet effect of von Holst but its apparently unitary nature contrasts with his original formulation of a compromise of two or more clocks. A notable feature of the timekeeper is that it is confined to the cycle-to-cycle timing of the rhythm. Within-cycle phasing of events is the duty of the first motor aspect. The timekeeping function, therefore, is probably not captured by the familiar neural concept of CPG. The first motor aspect provides a better fit to the traditional "CPG role." These conclusions agree with a recent suggestion (Lennard, 1985) that the neuronal underpinnings of the patterning of activity within a cycle are functionally distinct from the neuronal underpinnings of the cycle-to-cycle timing of locomotory activity.

REFERENCES

Cavagna, G. (1977). Elastic energy in muscle. *Exercise and Sport Science Review, 5,* 89-129.

Chatfield, C. (1975). *The analysis of time series: An introduction.* London: Chapman and Hall.

Grillner, S. (1975). Locomotion in vertebrates: Central mechanisms and reflex interaction. *Physiological Review, 55,* 247-304.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*, 347-356.

Holst, E. von (1937,1939/1973) *The behavioral physiology of animal and man.* Coral Gables, FL: University of Miami Press.

Lennard, P. R. (1985). Afferent perturbations during "monopodal" swimming movements in the turtle: Phase-dependent cutaneous modulation and proprioceptive resetting of the locomotor rhythm. *Journal of Neuroscience, 5*, 1434-1445.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pearson, K. G. (1976). The control of walking. *Scientific American, 235*, 72-86.

Rosenblum, L. D., & Turvey, M. T. (1988). Maintenance tendency in coordinated rhythmic movements: Relative fluctuations and phase. *Neuroscience, 27*, 289-300.

Ten Hoopen, M., & Reuver, H. A. (1967). Analysis of sequences of events with random displacements applied to biological systems. *Mathematical Bioscience, 1*, 599-617.

Turvey, M. T., Rosenblum, L. D., Schmidt, R. C., & Kugler, P. N. (1986). Fluctuations and phase symmetry in coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 564-583.

Turvey, M. T., Schmidt, R. C., Rosenblum, L. D., & Kugler, P. N. (1988). On the time allometry of coordinated rhythmic movements. *Journal of Theoretical Biology, 130*, 285-325.

Wing, A. M. (1980). The long and the short of timing in response sequences. In G. E. Stelmach & J. Requin, (Eds.), *Tutorials in motor behavior* (pp. 469-486). Amsterdam: North Holland Publishing Company.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14*, 5-12.

Yamanishi, J., Kawato, M., & Suzuki, R. (1979). Studies on human finger tapping neural networks by phase transition curves. *Biological Cybernetics, 33*, 199-208.

FOOTNOTES

[*] *Neuroscience*, 33, 1-10 (1989).

[†] Also University of Connecticut, Storrs.

[††] Now at the University of California at Riverside, Department of Psychology.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 243-249

An Analogy between the Mucosal Waves of the Vocal Folds and Wind Waves on Water

Richard S. McGowan

An analogy is made between the mucosal waves of the vocal folds, as represented by the two-mass model, and a model for the growth of wind waves on water. A layered-structure picture of the vocal folds helps in making the analogy, because this structure shares attributes of both situations. The simple analysis presented here suggests that the two-mass model may be modified to become a more histologically realistic, layered-structure model for mucosal wave dynamics.

INTRODUCTION

Waves, known as the mucosal waves of the vocal folds, can be observed on the surfaces of the vocal folds during phonation. This wave motion is a major part of the opening and closing of the vocal folds, particularly during the chest mode of phonation. The mucosal waves, observed under controlled conditions by Baer (Baer, 1975), can be seen under laryngoscopic examination. Wave fronts, or perturbations, on the surfaces of the folds move from the lower margins of the folds near the trachea to the upper margins of the folds near the vocal tract. The wave fronts extend along the length of each vocal fold from the posterior to the anterior portion of the fold.

One mechanical model of these waves is the two-mass model (Ishizaka & Flanagan, 1972; Ishizaka & Matsudaira, 1972) (see Figure 1). This model captures essential features of the air-solid interaction during phonation. (In this discussion and in the figures the air flow is horizontal along the x-direction, so the trachea is to the left and the vocal tract is to the right. The midline of the glottis is in the positive z-direction from one fold, and in the negative z-direction from the other fold.)

The masses are sprung to an underlying substrate so that they can vibrate up and down in the z-direction. There is also a spring between the masses, the coupling spring, which tends to pull the masses so that they have the same z-coordinate. When the coupling spring between the masses is compliant enough, mass 2 lags mass 1 by as much as a quarter cycle during a simulation of chest mode phonation - mimicking the action of the mucosal wave. The wave travels from left to right, or from mass 1 to mass 2 , in the coordinate system of Figure 1.

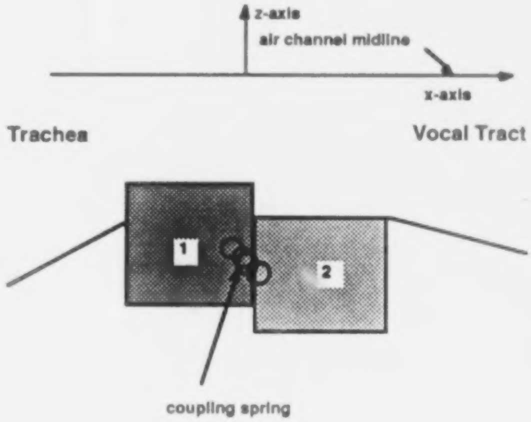


Figure 1.

This work was supported by NIH grants HD-1994, NS-13870, and NS-13617 to Haskins Laboratories. The author would like to thank Dr. Philip Rubin for comments on this manuscript and Dr. Carol Gracco for tutoring on histology.

An analogy to the the mucosal waves can be found in wind waves on water. The air motion along the water's surface can cause an instability, so that surface water wave motion ensues. Models have been invented to explain the initial stages of the growth of these waves from small perturbations. One of these models will be considered here in relation to the growth of the mucosal wave. To expedite the analogy between the two-mass model of the vocal folds and wind waves on water, the two-mass model is supplemented with a simplified layered-structure picture of the vocal folds, based on Hirano's studies (Hirano, 1981). The simplified layered-structure picture not only helps in the explanation of the mucosal waves, but it points to directions for future research in modeling the waves in a more realistic structure than provided by the two-mass model. While some of the underlying analysis used in this paper is mathematical in nature (Ishizaka & Matsudaira, 1972; McGowan, 1989), physical arguments can be used to convey the fundamentals of the analogy between the mucosal waves of the vocal folds and a model for wind waves on water.

WIND WAVES ON WATER

A consideration of other layered structures that show wave-like behavior when air passes over a surface is useful in understanding mucosal waves. One example is that of wind waves on water (Chandraesekhar, 1981; Lamb, 1945). This is a very simple layered structure consisting of two layers: air and water. Another, is structural instability resulting in vibration or flutter in the presence of air flow (Miles, 1956). Some common examples include a flag flapping in the breeze, or the destructive instability of aircraft panels that can occur during flight. In these examples there is the flow of air over a surface, which may or may not respond with wave motion or vibration, depending on its own mechanical properties, as well as those of the air. These are well studied areas in applied mechanics, and work in these areas can be studied for physical analogies. They all share the property that energy is transferred from the air to the surface to get wave action on the surface when the wind speed is great enough. Wind waves on water will be considered here because there is a model of wind wave growth on water that involves some of the same physics as that of the growth of the mucosal wave.

What is the instability that causes the growth of a wave from a small perturbation? When the water's elevation is perturbed above the flat, planar boundary by an amount ζ, there are the forces of gravity and surface tension to pull it back to the flat configuration (see Figure 2).

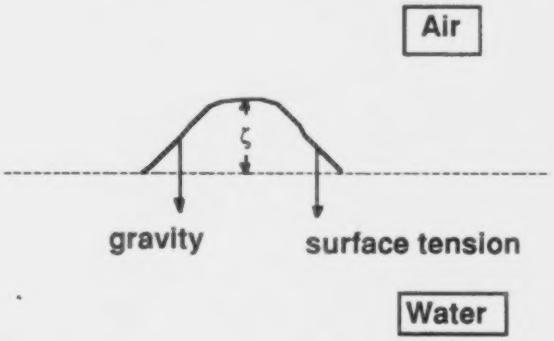


Figure 2.

It is assumed that there are always small perturbations to the base condition, so that there are spontaneous changes in the water's elevation. In equilibrium, with no perturbation, the air and water pressures are equal to P_0. If the pressures were not the same, there would be a net force at the surface of the water, and, hence, acceleration of the surface. But, when there is a perturbation to the surface elevation, the air pressure is perturbed by an amount p_a, and the water pressure by an amount p_w. These pressures in turn affect the surface elevation, ζ. Because elevation is a function of pressure, and pressure is a function of elevation, there is a feedback in the air-water system(see Figure 3). If, because of this feedback, the perturbation grows with respect to time, the configuration is said to be unstable and a wave may grow.

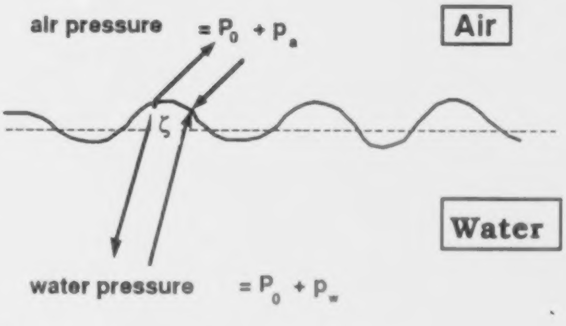


Figure 3.

To understand instability in this context, it is necessary to understand Bernoulli's relation, including the necessary conditions for its application. In the absence of friction and other energy loss mechanisms, in the absence of rotational motion of the air, and assuming the acceleration of the air can be neglected, the sum of pressure and kinetic pressure—total pressure—is constant. Kinetic pressure of air is the density of the air times the square of the air speed divided by two. Under these conditions of ideal flow, where Bernoulli's relation holds, the air moving at low speed is at a higher pressure than air at high speed: this is the Bernoulli effect. When a streamlined object is placed in a straight flow, the nose is under maximum pressure, because the flow is stagnant (see Figure 4). However, the pressure near the top of the object is slightly less than if there was no object in the flow. This is because the air accelerates over the top, giving a pressure drop by Bernoulli's relation (Lighthill, 1986). Given the necessary conditions for Bernoulli's relation in a region of flow, there is an inverse relation between pressure and flow speed.

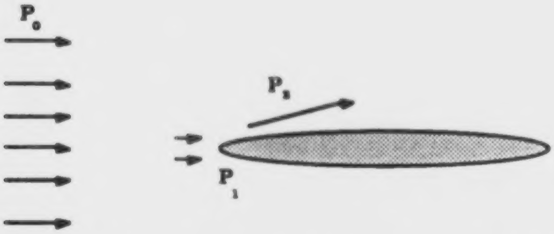


Figure 4. Total pressure is equal everywhere, so that the air pressures are related $P_2 < P_0 < P_1$.

At first the air flow resulting from the perturbations to the water's surface is assumed to follow Bernoulli's relation. That is, the perturbed water's surface is presumed streamlined and the flow of air ideal. (The quasi-steady approximation is used here, so the unsteadiness of the flow is neglected.) It can be seen that ideal flow is not the correct physical simplification in deriving the observed instability, and more realistic conditions must be incorporated to derive the instability.

To get a wave to grow with ideal flow conditions, it is necessary to counteract the restoring forces of gravity and surface tension. When the wind blows the air pressure decreases from what it was at rest as a result of the Bernoulli effect, and it is conceivable that the air pressure becomes so low so as to completely overcome the restoring forces (Miles, 1960). This is known as a static instability where, at least initially, the water is literally sucked into the air as if by a vacuum cleaner. Here there is an exponential growth in the elevation of the water. However, because the restoring forces of gravity and surface tension are so strong, the wind velocities required for this are much higher than those observed during wind wave production. (In the case of the flag, there is little restoring force perpendicular to the vertical plane, so this type of instability can get the flag to flap in the breeze.) Requiring that the restoring forces be overcome is too stringent. All that is needed is for energy to be fed from the air into the water at such a rate so as to overcome the any energy dissipation mechanisms. A negative resistance must be added to the system so that a different type of instability results: dynamic instability.

To derive a dynamic instability, the assumption of ideal flow of air must be abandoned. This can be seen by considering the following example. Suppose a sinusoidal perturbation is created on the surface of the water. If the air flow is symmetric over the maxima and minima, as it would be in the case of ideal flow, then both the air and water pressures are symmetric about the maxima and minima. This results in no net energy transfer from the air into the water, that is, there is no net power input from the air to the water. In the case of water-air surfaces, power input per unit area is calculated as the water pressure at the surface minus air pressure at the surface, which gives net pressure at the surface, times the vertical velocity of the surface. When the surface of the water is perturbed, there is vertical movement of the surface. Because of the symmetry of a sinusoidal perturbation, for each section going up with a given net pressure there is a section going down with the same speed and same net pressure, so that one product of net pressure difference times velocity will be the negative of the other. Therefore, there is no net power input from the air to the water. A feedback mechanism capable of breaking the symmetry provided by ideal flow is needed for a dynamic instability that can explain the formation of wind waves over water at a relatively low wind speed.

It is instructive to consider the first model of dynamic instability for wind waves on water. Despite the fact that it is not the correct model for the growth of wind waves on water (Miles, 1957; Lighthill, 1968), this model provides an analogy for the mucosal wave. Jeffreys (Jeffreys, 1925) proposed that the symmetry in air pressure between the windward and leeward side of a maximum is broken because of the phenomenon of flow separation. That is, vorticity is shed from the wave crests, which means that ideal flow no longer exists in all regions of the air flow, because the presence of vorticity means that rotational fluid motion is present. This symmetry breaking mechanism is known as Jeffreys' sheltering hypothesis. This is, for example, what happens to a blunt object, like a house, in wind, where vorticity is shed from the roof (see Figure 5). Because energy is transferred into the vorticity, the total pressure on the windward side of the house is greater than that on the leeward side, and the stronger the vorticity the greater the difference in total pressure. Further, because the air speed is zero along all surfaces of the house, the air pressure on the windward side is greater than that on the leeward side.

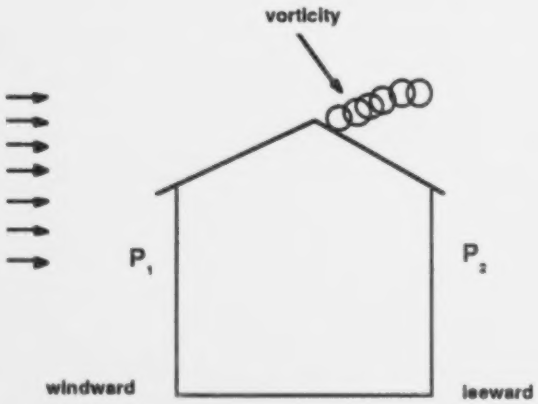


Figure 5. Total pressure on the windward side > total pressure on the leeward side, so that the air pressures are related $P_1 > P_2$.

With the abandonment of ideal flow in the case of the air-water interface, there is a difference in the total pressure on the windward side and the leeward side of a crest. For water waves traveling in the same direction as the wind, the sides moving downward are also windward and, hence, under greater air pressure than the leeward sides which are moving upward. This means that more energy is fed into the water in the downward motion than is taken away in the upward motion. The net energy transfer is into the water, if the initial sinusoidal perturbation represents a wave traveling in the same direction as the wind. Wave growth ensues if the energy gained in a cycle is greater than that lost by dissipation mechanisms. This is a dynamic instability, where there is exponential growth of oscillation rather than the pure exponential growth seen for static instability.

The relevance of the above considerations to the two-mass model of the vocal folds is that there is vorticity formation at the glottal entrance and exit. This allows energy to be fed from the air into the vocal folds in the event the surfaces of the folds at the glottal entrance and exit are out of phase. Before the details of the analogy are spelled out, the layered-structure picture of the model folds is described.

LAYERED STRUCTURE AND ITS RELATION TO THE TWO-MASS MODEL

To draw an analogy between Jeffreys' model for the growth of wind waves on water and the mucosal waves as described by the two-mass model, some of the essential histological features of the vocal folds are used to build the layered-structure picture of the vocal folds. In particular, the description of the layered structure of the folds near the edges as provided by Hirano (Hirano, 1981) is used (see Figure 6). Consider a single fold. The epithelium is in contact with the air, which is flowing in the x-direction. Below this (in the sense of away from the mid-line) is the superficial layer of the mucosa, otherwise known as Reinke's space. The intermediate and deep layers are ligamentous and muscular layers, and these are grouped together in the simple layered-structure picture presented here. Note that the two-dimensional aspect of the two-mass model is being followed in the layered-structure description of the folds, so variations in the anterior-posterior dimension are not allowed.

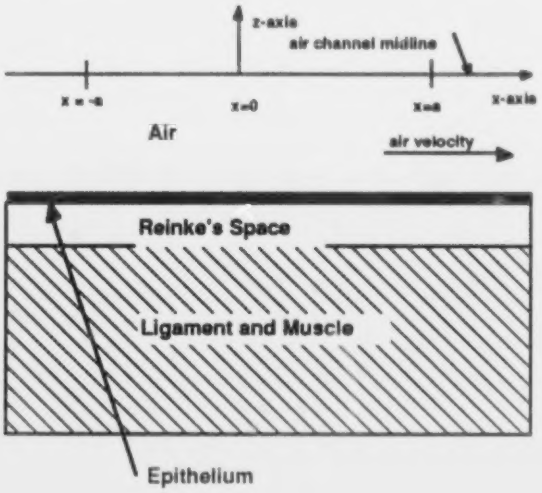


Figure 6.

The epithelium is a thin, elastic sheet. The epithelium is assumed to be relatively elastic because of the tight interconnections between the cells provided by the desmosomes (Junqueira, Carneiro, & Long, 1986). This elasticity provides a restoring force when the epithelium is stretched and perturbed similar to the way a drumhead behaves. It is important to note that the shorter the length scale, or wavelength of a perturbation, the stronger the restoring force. [This is only a linear analysis, the change in tension with amplitude of vibration is not be taken into account (Titze & Durham, 1987). Further, bending stiffness could be taken into account in a more detailed analysis (Bickley, 1987).]

In contrast with the epithelium, the material of Reinke's space behaves more like an inelastic fluid. Reinke's space is an amorphous suspension of collagenous fibers (Junqueira, Carneiro, & Long, 1986). The fact that these fibers are not connected together means that the elasticity of the fibers do not act to make Reinke's space elastic. The Reinke's space pressure will also be affected by the elevation of the epithelium.

Finally, the ligamentous and muscle layers constitute a single layer in this simplified picture. This layer reacts to perturbations and stresses from the other layers, and the reaction is parametrized as an impedance seen from Reinke's space. For a thin Reinke's space, the mass, stiffness, and damping add to corresponding quantities of the epithelium, so that changes in mechanical quantities in the underlying ligament-muscle layer have a direct effect on the mechanics of the epithelium.

The only changes from the two-mass model in adopting the layered structure picture are to replace the masses in the two-mass model with the layered structure, and to replace the springs, which provide the restoring forces in the two-mass model with the tension in the epithelium and the underlying ligament-muscle stiffness. The laws governing the air flow are left unaltered. The air flow obeys Bernoulli's relation, except at the entrance and exit of the glottis, where there can be change in the total pressure. The total pressure decreases from the tracheal region to the glottal region because of vorticity at the glottal entrance, and the total pressure decreases from the glottal region to the vocal tract because of vorticity formation at the glottal exit. However, Bernoulli's relation holds locally within the trachea, glottis, and vocal tract. Feedback exists here because the amount of vorticity depends on the the elevation of the epithelium within the glottal region and the amount of vorticity, in turn, determines the air pressure at the epithelial surface. This feedback can provide for an instability for wave growth on the epithelium.

INSTABILITY OF THE LAYERED STRUCTURE

With the layered-structure picture, the stability of the epithelial surface is analyzed in essentially the same way that Ishizaka and Matsudaira analyzed the stability of the two-mass model (Ishizaka & Matsudaira, 1972; McGowan, 1989). To study stability, it is necessary to consider what occurs with air flowing at velocity U and given that the epithelium has a perturbation elevation, ζ, when the equilibrium configuration is one with plane boundaries between the layers and no air flowing. In the perturbed state, both the air pressure and Reinke's space pressure are perturbed, as is the ligament-muscle layer. Because the pressures in each the air and Reinke's space depends on the elevation, ζ, there is feedback occurring. For instance, as the epithelium encroaches on the air channel in the form of a bump, there must be a slight increase in air speed over the bump. This, in turn, means a slight decrease in pressure by Bernoulli's relation, and this tends to make the bump grow in amplitude. If it were not for the combined restoring force of the elastic epithelium and the underlying ligament-muscle layer, this bump would just keep on growing.

It must be noted that the Bernoulli effect is amplified when the air is channeled as it is in the glottis. Not only is the air speed greater because of the constriction, but it can be shown mathematically (Ishizaka & Matsudaira, 1972; McGowan, 1989) that the boundary conditions provided by the channeling of the air greatly amplifies the importance of the air flow and, in particular, the Bernoulli effect. Because the Bernoulli effect acts against the restoring forces, the effective restoring force of the epithelium can be greatly reduced with increased air flow and the narrowing of the glottal channel. Note that it is the combined effects of the epithelial tension, underlying ligament-muscle stiffness, and the Bernoulli force that determines the total effective restoring force. It is not any one effect determining the vibratory characteristics of the folds in small amplitude vibration.

Because of the large Bernoulli effect, it is not surprising that static instability is possible for the epithelium. Recall that for wind waves on water that static instability results when the reduction of pressure due to the Bernoulli effect overcomes the restoring forces of gravity and surface tension. The conditions under which this would occur in the case of the vocal folds are a loose epithelium and loose vocalis, perhaps meaning a relaxed cricothyroid muscle, as well as closely approximated folds and a fairly healthy airflow, both to enhance the Bernoulli effect. It may be that the closing phase of creaky voice can be understood as having conditions near static instability for a good part of the closing phase.

However, this severe static instability cannot be relied on to produce the mucosal wave in chest voice. The Bernoulli effect may lessen the effective epithelial tension, but it does not overcome the restoring forces. As in the case of wind waves on water, vorticity must be brought into the picture. Ideal flow conditions giving Bernoulli's relation do not hold at the entrance and exit of the glottis because of the vorticity provides an additional feedback mechanism.

This feedback is a little more complex than Jeffreys' sheltering mechanism for wind waves on water because volume velocity, in addition to total pressure, is changing with the vorticity formation in the case of the vocal folds. Consider the glottis to extend from -a to a along the length of the of the epithelium in the x-direction. Figure 6 shows a schematic of the layered-structure picture of one of the folds with one-half the air channel. Mathematical analyses show that the feedback mechanism for the vocal folds only works on the part of the epithelium closest to the trachea, the lower margins, say, between -a and 0. ('Lower margins' in the coordinate system here are the parts of the folds furthest to the left.) The parts of the epithelium closest to the vocal tract, the upper margins, between 0 and a, are relatively unaffected by this mechanism. If the lower margins of the epithelium are moved toward the midline, the consequence of increased vorticity formation at the glottal entrance is an air pressure deficit at the lower margins. With increased vorticity formation at the glottal entrance comes a greater difference in total pressure between the trachea and the glottis, and, hence, a smaller total pressure for the glottal region. As long as the air speed is maintained sufficiently over the lower margins, this will mean a pressure drop in the air pressure over the lower margins. If the upper margins of the epithelium are raised, there is a pressure excess at the lower margins of the folds. With greater vorticity formation at the glottal exit comes a greater difference in total pressure between the glottal region and the vocal tract. This means that there is less difference in total pressure between the trachea and the glottis, because the difference in total pressure between the trachea and vocal tract is constant. So, there is an increase in total pressure in the glottal region with elevated upper margins. As long as air speed does not increase too much over the lower margins, there must be a pressure increase over the lower margins.

Given these facts, if the upper margins of the epithelium are somewhere around a quarter cycle behind the phase of the lower margins, then the following events occur during a cycle. While the lower margins of the epithelium are moving toward the midline, the upper margins are away from the midline, giving a pressure deficit on the lower margins. There is a pressure excess on the lower margins while the lower margins of the epithelium are moving from the midline because the upper margins are high. Therefore, this feedback is such that the pressure on the lower margins is greater during its downward motion than on its upward motion, and there is net power input from the air into the epithelium. If the power is great enough to overcome the rate of dissipation by damping, then surface waves grow (Stevens, 1977). This is the scenario for waves of wavelength of around 4a traveling in the direction of flow, because such waves provide the correct phasing between the upper and lower margins of the folds.

As in the case of Jeffreys' sheltering hypothesis, vorticity is used to break the symmetry in the pressures on the upstroke and downstroke of a surface. This is essentially a feedback mechanism, because the strength of the vorticity depends on the height of the surface.

CONSEQUENCES AND FUTURE RESEARCH

The analogy between explanations of wind waves on water and the mucosal waves of the vocal folds adds to the understanding of the physics of phonation. For instability to result, the energy input by a feedback mechanism must overcome any dissipation of energy. Energy dissipated in each cycle increases with frequency and, hence, restoring force. So, air flow helps to create the dynamic instability in two ways: It provides for vorticity formation and it reduces the effective restoring force of the folds through the Bernoulli effect in the glottal constriction, thus reducing the rate of dissipation.

There must be a phase difference between the upper and lower margins of the folds for the kind of instability described here to occur. This puts an upper limit on the wavelength of the mucosal wave, because for long wavelengths, the upper and lower parts are nearly in phase. However, there is one more necessary condition for this instability: The energy input by the special feedback mechanism must be greater than the energy lost by dissipation. Energy dissipation increases with decreasing wavelength. The two necessary conditions for dynamic instability have opposite wavelength dependence: phasing requiring shorter wavelengths and that of keeping dissipation bounded requiring longer wavelengths. Under certain conditions, such as a tense cricothyroid, these requirements may be incompatible, so that another type of instability must be sought to explain wave growth. The mode of vibration when the two conditions become incompatible is one-mass vibration seen in falsetto voice.

The use of a layered-structure picture of the vocal folds has expedited the analogy between the mucosal wave of the vocal folds and wind waves on water. It is hoped that the feasibility of a layered-structure approach to vocal-fold mechanics has been demonstrated in this paper. More explanation of the voice modes using this approach is not possible without some added, important factors, such as the third, anterior-posterior dimension and nonlinear interaction. Numerical computation will eventually be needed for further progress.

REFERENCES

Baer, T. (1975). *Investigation of phonation using excised larynxes.* Doctoral dissertation, M.I.T.

Bickley, C. (1987). Modeling the acoustic characteristics of children's speech: Fundamental frequency. In *Proceedings of the Eleventh International Congress of Phonetic Sciences.* Tallinn, Estonia, U. S. S. R., .

Chandraesekhar, S. (1981). *Hydrodynamic and hydromagnetic stability* (pp. 481-92). New York: Dover.

Hirano, M. (1981). *The clinical examination of voice* (pp. 4-7). Vienna: Springer-Verlag.

Ishizaka, K., & Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *The Bell System Technical Journal, 51,* 1233-1268.

Ishizaka, K., & Matsudaira, M. (1972). *Fluid mechanical considerations of vocal cord vibration* (Monograph 8). Santa Barbara, CA: Speech Communications Laboratory, Inc.

Jeffreys, H. (1925). On the formation of water waves by wind. *Proceedings of the Royal Society of London, 107,* 189-206.

Junqueira, L. C., Carneiro, J., & Long, J. A. (1986). *Basic histology* (p. 72). Norwalk, CT: Appleton-Century-Crofts.

Lamb H. (1945). *Hydrodynamics* (pp. 461-62). New York: Dover.

Lighthill, M. J. (1968). Physical interpretation of the mathematical theory of wave generation by wind. *Journal of Fluid Mechanics, 14,* 385-397.

Lighthill M. J. (1986). *An informal introduction to fluid mechanics* (p. 171). Oxford, England: Clarendon Press.

McGowan R. S. (1989). Phonation form a continuum mechanics point of view. (To appear in the *Proceedings of the Vocal Fold Physiology Conference* held in Stockholm, Sweden)

Miles J. W. (1956). On the aerodynamic instability of thin panels. *Journal of Aeronautical Sciences, 23,* 771-780.

Miles J. W. (1960). On the generation of surface waves by shear flows, part 3. Kelvin-Helmholtz instability. *Journal of Fluid Mechanics, 6,* 583-598.

Miles J. W. (1957). On the generation of surface waves by shear flows. *Journal of Fluid Mechanics, 3,* 185-204.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx models. *Phonetica, 34,* 264-279.

Titze I. R., & Durham P. L. (1987). Passive mechanisms influencing the fundamental frequency control. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Laryngeal function in phonation and respiration.* Boston: College-Hill Press.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 250-255

Laryngeal Control in the Plosives of Standard Thai*

Arthur S. Abramson†

Much work has been done on the relevance of the timing of glottal pulsing relative to the occurrence of consonantal closure for phonemic distinctions between homorganic stops. In this paper, I have assembled old and new work on the plosives of Standard Thai, including the addition of data on the palatal affricates. In word-initial position, in both production and perception, the traditionally labeled voiced, voiceless unaspirated, and voiceless aspirated stop consonants are well separated by voice onset time (VOT). The two palatal affricates are likewise so distinguished in production, although perceptual experiments have not yet been run for them. In word-final position, where there are no contrasts of voicing and aspiration, the only plosives that occur are the stops; they show no voicing in their closures. A limited sampling of running speech indicates that, with some reduction of range, the distinguishing mechanism is quite robust. Temporal control of the laryngeal source, then, is sufficient to account for the three-way voicing distinction of Standard Thai.

BACKGROUND

In common with other languages of the Tai family, Standard Thai has a system of plosive consonants that is interesting for two reasons: (1) its three-way voicing contrast and (2) its paradigmatic gaps. This can be seen in the intersecting rows and columns of Table 1. With some minor disagreements, this table reflects the organization given or implied by various scholars (e.g., Abramson, 1962; Harris, 1972; Noss, 1964). The two bottom rows are usually said to be distinguished by the feature of aspiration, even though the traditional phonetic meaning of the term is not properly applicable to affricates, as in the case of the palatal plosives of Thai. That is, the term is used here in an abstract phonological sense; nevertheless, as will be shown later, the palatals agree with the simple oral stops in that we can separate them by a measure of the relative timing of the voice source. Finally, both the palatals and the velars have "holes in the pattern" in the voiced series.

Table 1. *Paradigmatic array of the oral plosives of Standard Thai.*

	Bilabial	**Dental**	**Palatal**[a]	**Velar**
Voiced	b	d		
Voiceless unaspirated	p	t	c	k
Voiceless aspirated	ph	th	ch	kh

[a] The 1989 revision of the chart of the International Phonetic Alphabet has eliminated the old separate palato-alveolar column.

The palatals appear only in syllable-initial position; furthermore, the three-way contrast of the three rows of the table is relevant only in that position. Most linguists working on Thai take the simple oral stops in final position to be /p t k/.

Some years ago, Leigh Lisker and I wondered whether a single laryngeal mechanism might underlie distinctions between homorganic consonants commonly described, for so many languages, as depending upon such features as voicing, aspiration, tensity, and perhaps others. Such a mechanism posited by us was one of laryngeal timing. That is, we asked whether a "simple" difference in the timing of the turning on and off of the laryngeal voice-source relative to the

---

This work was supported by Grant HD-01994 from the U.S. National Institute of Child Health and Human Development to Haskins Laboratories. I am grateful to Professor Leigh Lisker for his helpful comments.

moment of consonantal occlusion or constriction would serve to differentiate the phonemes in question. We limited ourselves in that study (Lisker & Abramson, 1964) to stop consonants and to word-initial position to take advantage of what seemed to be the maximum number of such contrasts in a wide variety of languages.

Limited thus to initial position, our measurement of laryngeal timing became one of voice onset time (VOT), which was defined as the time between the onset of voicing and the release of the stop. With zero time assigned to the moment of release, voicing onset before the release was stated as a negative value and onset after the release as a positive value. Using sound spectrograms, we obtained data from a number of languages with a two-way contrast, such as English, a three-way contrast, such as Thai , and a four-way contrast, such as Hindi. By and large, our VOT values did a good job of separating the categories of the 11 languages examined.[1] Experiments with synthetic speech confirmed the perceptual efficacy of VOT for a representative sampling of languages (Abramson & Lisker, 1965, 1970; Lisker & Abramson, 1970). All this led to our arguing for the relevance of voice timing to questions concerning the validity of distinctive features in phonology (Lisker & Abramson, 1971).

Among the languages successfully handled by the concept of VOT is Standard Thai. As a token of my great respect and admiration for my old friend Mayuri Sukwiwat, I have pulled together old data on the plosives of the language from various publications and combined them with new data, including measurements of the previously unanalyzed palatals.

THAI DATA

Citation forms

Not only in the laboratory of the experimental phonetician, but also in linguistic fieldwork, it has been customary and very helpful to build up a data-base with short expressions, commonly citation forms of words, uttered by native speakers of the language under investigation. Although we know that in everyday speech communication, such words are much more likely to be embedded in running speech that varies in tempo, emphasis, and intonation, we nevertheless find citation forms to afford a convenient point of departure for further work. In any event, one-word sentences do occur from time to time. One might argue that, starting with the first learning of our native language in early childhood, we always have citation forms as one mode of speech. Perhaps they come closest to the underlying abstract lexical entries posited by the phonologist. Of course, a complete phonological account would provide rules for the contextually determined variations of these forms in running speech.

The data in Table 2 were obtained from wideband spectrograms of the productions of three native speakers of Thai. At that time (Lisker & Abramson, 1964), we rounded off all values to the nearest 5 milliseconds, as an estimate of our precision; this is reflected in the ranges. All tokens of /b/ and /d/ show negative values of VOT starting well before the release. The voiceless inaspirates /p t k/ show short voicing lags after the release, while the voiceless aspirates /ph th kh/ show long lags. That is, the temporal dimension serves well to separate the categories. In spite of the paradigmatic gap where there could be a voiced velar stop, there is no intrusion of /k/ values into that space.

Table 2. *Three speakers' voice onset times (in ms) for the oral stops of Thai in isolated words: means, ranges, and numbers of tokens. (From Lisker & Abramson, 1964)*

		Labial	
	/b/	/p/	/ph/
M	–97	6	64
R	–165:–40	0:20	25:100
N	31	32	33
		Dental	
	/d/	/t/	/th/
M	–78	9	65
R	–165:–40	0:25	25:125
N	33	33	33
		Velar	
		/k/	/kh/
M		25	100
R		0:40	50:155
N		32	38

New data for the present study were obtained from a recently available speaker, CT, whose production of the three-way contrast is exemplified for the dental stops in the waveforms of Figure 1. I have chosen to take advantage of the greater precision of an elaborate waveform-editing computer program at Haskins Laboratories, rather than use spectrograms again.

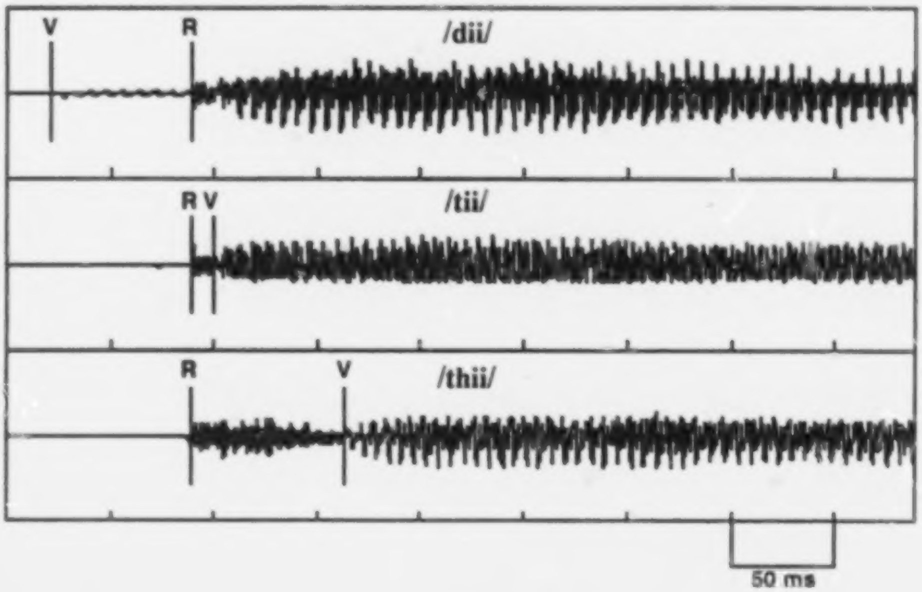


Figure 1. **Waveforms showing VOT differences in Thai dental stops. R=release of oral closure. V=onset of voicing (glottal pulsing).**

The moment of release of the stop, marked R, is generally easy to pick out. Voice onset is marked V. Let us look more closely at the waveforms of Figure 1. The releases of the initial stops of the three words have been aligned in time. The final portions of the final long vowels have been trimmed off to make for a symmetrical display. For /dii/, the quasi-periodic pulsing V–R occurs during the closure before the release and so takes a negative VOT value, i.e., a voicing lead of –68 ms. In both /tii/ and /thii/, the interval R–V at the beginning of the vocalic span is filled with aperiodic energy resulting from a turbulent airflow; that is, they have positive values of VOT. Thus, in these examples, /tii/ has a voicing lag of 10 ms and /thii/, 74 ms. The VOT values in Table 3 are, then, for all the monosyllabic utterances analyzed, the durations of the intervals between R and V. This time, I have given standard deviations rather than ranges. (The raw data from 1964 are no longer available, so I cannot calculate standard deviations for Table 2.) New in this table and, I believe, in the literature, are values for the palatal affricates. The mean values and the small standard deviations show good separation of /c/ and /ch/, with no intrusion into the space for a potential palatal with voicing lead. Some work has been done on the intersection of phonemic tones with the voicing states of Thai plosives (Gandour, 1974) and the involvement of these states with perturbations of fundamental frequency in the historical emergence of tones in the language (e.g., Erickson, 1975).

Table 3. *CT's VOT values (in ms) for the plosives of Standard Thai: Means, standard deviations, and numbers of tokens.*

		Labial	
	/b/	/p/	/ph/
M	–85	10	95
SD	–32	3	13
N	10	10	10
		Dental	
	/d/	/t/	/th/
M	–87	11	85
SD	–18	4	20
N	10	10	10
		Palatal	
		/c/	/ch/
M		38	124
SD		11	31
N		10	10
		Velar	
		/k/	/kh/
M		24	104
SD		8	20
N		10	10

As I intimated at the beginning of this paper, something has to be said to justify the feature aspiration in Table 1 for the palatals. The waveforms in Figure 2 will help. Comparing the latter with the waveforms of the voiceless stops in Figure 1 reveals that for both sets the interval R–V is filled with turbulence; however, the turbulence for the affricates is obviously more intense and of longer duration than that of the stops relative to the following voice-excited remainder of the vowel. This difference arises from the fact that the turbulence of the affricates is local; it is generated, with the glottis open, in the narrow channel of the lingual constriction of the fricative portion of the articulation. Thus, it is not aspiration in the usual sense of the term. Note, by the way, that such an account implies an additional difference between /c/ and /ch/, namely, an articulatory one; to allow through *local turbulence* for a longer voicing lag in /ch/, it is necessary to maintain the frication of this plosive for a longer time. Such an inference, I hasten to add, remains to be verified by direct examination of articulation, perhaps through dynamic palatography. (This means that the language has not exploited another phonetic possibility, aspiration after the relaxation of the palatal constriction.) In the voiceless stops,on the other hand, aside from the transient noise at the moment of release itself, most of the turbulence is caused by an aperiodic source at the open glottis exciting the relatively open, unconstricted supraglottal cavities suitable for the articulation of vowels.[2] This is true aspiration as the term is normally used, even though some small portion of the turbulence is no doubt contributed by friction along the walls of the tract. In any event, the data presented here lend support to the handling of all the Thai plosives in the same way. The abstract phonologist may see some virtue in continuing to use the word "aspiration" for both kinds of voicing lag.

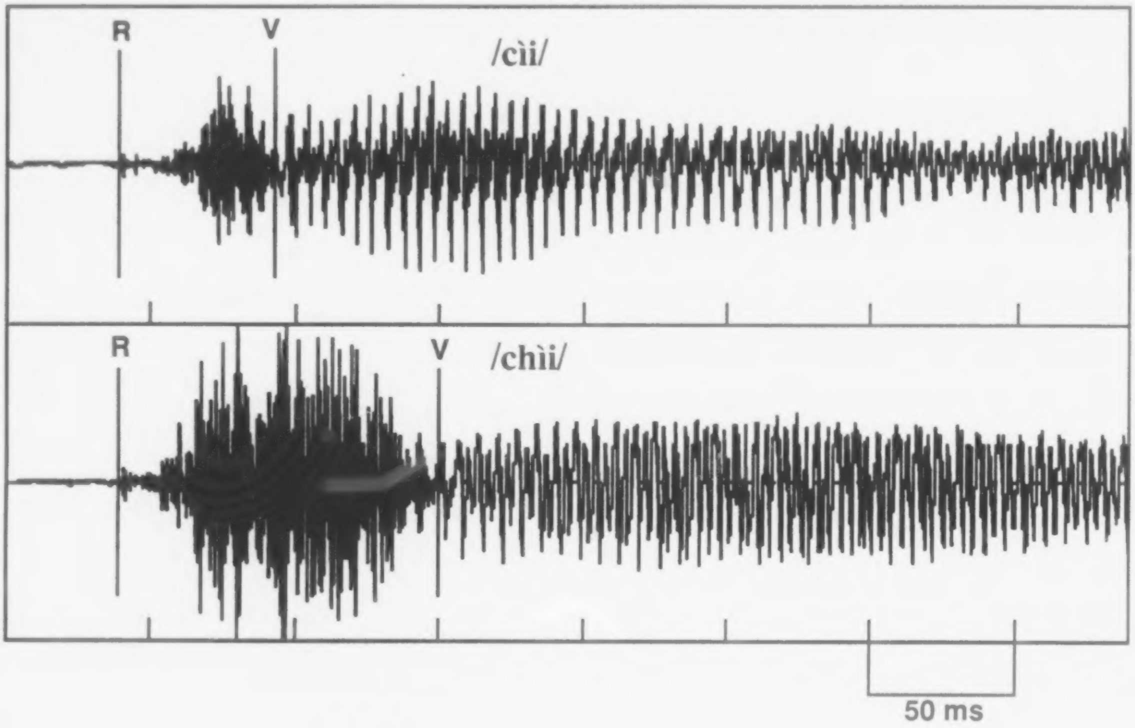


Figure 2. Waveforms showing VOT differences in Thai palatal affricates. R=release of oral closure. V=onset of voicing (glottal pulsing). R—V is also the duration of the frication.

Perceptual validation

Not only our original measurements (Lisker & Abramson, 1964) but also a perceptual study with manipulated and unmanipulated natural English speech with native speakers of a few languages, including Thai (Lotz, Abramson, Gerstman, Ingemann, & Nemser, 1960), led us to hypothesize that the complex acoustic consequences of systematic shifts in laryngeal timing are perceptually relevant. Producing small increments of VOT from -150ms to +150ms for labial, dental, and velar stops on a formant synthesizer, we produced synthetic syllables that varied in a number of ways: (1) time of onset of pseudo-glottal pulsing and thus in the spectral nature of the voicing, which might start as a low-frequency band during oral closure or as periodic excitation of vocalic formants, (2) aspiration noise during voicing lag, and (3) attenuation of the first formant of the vowel during voicing lag. We had native speakers write words in Thai script as labels for all these stimuli. The results strongly supported the hypothesis (Abramson & Lisker, 1965, 1970; Lisker & Abramson, 1970).

Running speech

Perhaps the ultimate test of the validity of phonetic features found to underlie phonemic distinctions in citation forms is the power of such features to distinguish phonological categories in casual running speech too. There is, unfortunately, a difficult procedural problem. For a given speaker, it is very hard to gather a large, contextually balanced sampling of measurable tokens of plosives in running speech. If the speech is truly spontaneous and unrehearsed, there is no way to keep the speaker from varying his overall prosody and embedding his plosives in a random assortment of phonetic contexts.

In spite of these problems, I have measured VOT values for a number of plosives from about 10 minutes of a description of Thai holidays recorded for another purpose by CT, the new speaker of this study. The data are given in Table 4. The numbers of utterances for the set of plosives are not only uneven but also small, so I have entered them all rather than give averages. In spite of some compression of ranges, especially for the aspirates, it can be seen that the phonetic distinctions are maintained. The results are very similar to those obtained long ago by Lisker and me (1964) for one other speaker. In both sets of data, I could not measure many additional tokens of /b d/ in which glottal pulsing continued unbroken from a preceding voiced environment. Of course, inseparable as they are for purposes of measurement from preceding segments, these items do, nevertheless, fall into place as standing in clear contrast, through laryngeal timing, with the other two categories. I plan soon to test the robustness of VOT for perception in running speech.

Table 4. *CT's VOT values (in ms) for the plosives of Standard Thai in a small sample of running speech.*

	Labial	
/b/	/p/	/ph/
–80	10	51
–90	5	74
–80	3	
	Dental	
/d/	/t/	/th/
–27	12	79
–109	9	34
	11	
	13	
	Palatal	
	/c/	/ch/
	25	78
	37	94
		91
	Velar	
	/k/	/kh/
	18	53
	21	69
	29	100
		63
		90
		44

Another aspect of running speech, one that pertains to isolated words as well, is the voicing state of closures of word-final stops. (It will be recalled that the palatals do not occur here.) I explored this question in another paper (Abramson, 1972). The overwhelming finding was that, whatever the following environment, the closures of word-final Thai stop consonants are not voiced. A scattering of additional data from the speech of CT of this paper is in complete agreement with that conclusion. Thus, given the suspension of the three-way contrast in this position, it is best to align the word-final oral stops with initial /p t k/.

CONCLUSION

The efficacy of a feature of laryngeal timing as the phonetic basis of the three-way distinction of voicing and aspiration in the plosives of Standard Thai has been established. With a phonologically required focus on word-initial position, this mechanism is conveniently called voice onset time (VOT). Various experiments have shown the perceptual relevance of this feature for stops but not yet for the palatal affricates. Acoustic underpinning has been provided for the assignment of final stop consonants to the category of voiceless inaspirates. Limited work with running speech suggests the robustness of the mechanism, although perceptual work remains to be done.

Of theoretic interest for Thai and other languages in which laryngeal timing is critical for consonantal distinctions, is the observation that a single physiological mechanism appears to underlie a rather complex acoustic output. That is, a number of acoustic features of this output might separately serve as acoustic cues to the phonemic distinctions, yet all of them seem to be linked under the control of a single mechanism.

REFERENCES

Abramson, A. S. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments. (Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20 [*International Journal of American Linguistics, 28*, No. 2, Part III]). Bloomington.

Abramson, A. S. (1972). Word-final stops in Thai. In J. G. Harris & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 1-7). Bangkok: Central Institute of English Language.

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34*, 295-303.

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Commins (Ed.), *Proceedings of the Fifth International Congress on Acoustics. Ia* (paper A51). Liège: G. Thone.

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569-573). Prague: Academia.

Erickson, D. (1975). Phonetic implications for an historical account of tonogenesis in Thai. In J.G. Harris & J.R. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 100-111). Bangkok: Central Institute of English Language.

Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of Phonetics, 2*, 337-350. In J. G. Harris & R. B. Noss (Eds.) *Tai phonetics and phonology* (pp. 8-22). Bangkok: Central Institute of English Language.

Harris, J.G. (1972). Phonetic notes on some Siamese consonants. In J. G. Harris & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 8—22). Bangkok: Central Institute of English Language.

Kim, C. W. (1970). A theory of aspiration. *Phonetica, 21*, 107-116.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language, 47*, 767-785.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. J. (1960). The perception of English stops by speakers of English, Spanish, Hungarian and Thai: A tape-cutting experiment. *Language and Speech, 3*, 71-77.

Noss, R. B. (1964). *Thai reference grammar*. Washington, DC: Foreign Service Institute, Department of State.

FOOTNOTES

[*] In Supanee Tiancharoen (Ed.), *Festschrift for Mrs. Mayuri Sukwiwatt. PASAA, 19*, 85-93 (1989).

[†] Also University of Connecticut, Storrs.

[1] To separate all four categories of Hindi and Marathi on the one hand and all three of Korean on the other, VOT must be supplemented in each case by one more laryngeal feature.

[2] Jimmy G. Harris (1972) states that the voiceless inaspirates of Thai are produced with a closed glottis. Of course, one way of cutting off glottal pulsing is through tight adduction of the vocal folds. If he is right, it must nevertheless be the case that the glottal stop occurs during the oral closure, because, as can be seen in the waveforms here or in spectrograms, the glottis must be open during the voicing lag to yield the observed turbulence.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 256-268

Distinctive Vowel Length: Duration vs. Spectrum in Thai*

Arthur S. Abramson† and Nianqi Ren†

Many languages are described as having a phonological distinction of length in vowels or consonants or even both. If the term is to be taken literally, we would expect to find that the underlying mechanism is control of the relative durations of the articulations for the phonemes in contrast. In languages with short and long vowels, however, it is often observed that short and long counterparts have somewhat different vowel qualities. Previous work on the vowels of Central Thai has shown that relative duration is a sufficient acoustic cue for the length distinction. Since, however, short vowels in portions of the vowel space tend to be more open, we have sought to learn how much of the perceptual burden is borne by relative duration, the major physical correlate of length, and how much by the spectrum (formant pattern), the major physical correlate of vowel quality. We have incrementally lengthened original short vowels and shortened original long vowels in minimal pairs of words embedded in sentences, thus preserving the spectral differences. The stimuli were played to native speakers for identification. The results show that the dominant cue is relative duration; however, for all the vowel pairs, the category boundary is influenced by spectral differences with, perhaps, some effect of the timing of the context as well.

1. INTRODUCTION

As with all phonological distinctions, so also with the feature of vowel length the question arises as to the likelihood that the distinction rests on a single phonetic feature. On the face of it, distinctive length seems to be little more than a matter of relative timing. Thus, in a given context, members of a "long" category are longer than their counterparts in the "short" category. For this seemingly straightforward situation, Catford, in his remarkable new textbook (1988, pp. 184–186), gives short shrift to the matter of learning to make long segments by emphasizing the need just to hold an articulatory configuration longer. We are thinking here, of course, of the prolongation of segments under voluntary control, not intrinsic length correlated, for example, with close versus open vowels or stops versus glides (Lehiste, 1970).

If the acoustic outputs of the putative length distinction could be shown to be differentiated only by relative duration, all phonologists would surely not hesitate to posit a distinctive feature of length. Indeed, the occasional, perhaps contextually predictable, presence of minor concomitant acoustic features, some of them audible, might not worry these analysts. Lehiste (1970, pp. 30–33) tells us that in "quantity" languages some differences in the phonetic quality of long and short vowels can be observed, although these languages differ in how much length and quality are correlated. She illustrates the matter with plots of the first and second formants for the

---

This work was supported by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories. It is an expanded version of an oral paper given at the 63rd Annual Meeting of the Linguistic Society of America, New Orleans, Louisiana, December 27–30, 1988. The first author used much of the material, together with other work, in "The Phonetics of Distinctive Length," a talk given by invitation at the International Conference on Linguistic Approaches to Phonetics in Honor of Professor J.C. Catford, Ann Arbor, Michigan, May 1–3, 1989. He wishes to thank the National Research Council of Thailand and the Department of Linguistics, Chulalongkorn University for their hospitality and help, and The University of Connecticut for the sabbatical leave in 1988 that made obtaining the data in Thailand possible. Our patient and cheerful Thai speaker was Mr. Akaravudh Osathanukrah.

vowels of two languages with distinctive vowel length, Czech and Serbo-Croatian. It turns out, for example, with regard to placement in the acoustic space, that in Czech "the difference between long and short vowels is greatest for /i/ and /u/, whereas in Serbo-Croatian the greatest difference is observed in the mid vowels /e/ and /o/" (p. 32). Even in a monograph dedicated to the subject of distinctive vowel length, Durand (1946) assigns great importance to features that go together with length in a number of languages. Indeed, if vowel quality goes hand in hand with vowel length (Catford, 1977, pp. 195–199; Garnes, 1973; Straka, 1959), it may be decided that quality is paramount, or "tensity" may be invoked to "explain" the correlation between length and quality. Related to this is the general understanding that in Standard Italian and other languages with distinctively short and long *consonants*, features other than just relative duration of closure or constriction might be observed. We see, for example, that in Italian *fato* [faːto] 'fate' and *fatto* [fatːo] 'done,' what is generally accepted as distinctive consonant length entails a reciprocal difference in vowel length. Is this an entirely redundant rhythmic feature? Are we not entitled to ask whether perceptual identification of these words is determined by two acoustic cues?

For the historical linguist, the existence at a stage of the history of a language of a feature of distinctive segment length linked to other features, even if the latter can be shown to be entailed by length, offers the possibility of diachronic sound change. If conditions are ripe, one of the concomitant features could take over the communicative function of the old primary feature, resulting in a change in the system of phonemes. For example, the Indo-European feature of distinctive vowel length is said to persist in a number of Germanic languages, among others. In the southern dialect of one of them, Swedish, experimental work (Hadding-Koch & Abramson, 1964) has shown that a difference in quality correlated with length has seemingly taken over the distinctive function in at least part of the vowel system. That is, for one pair of "short–long" vowels, acoustic cues to the distinction reside in the formant patterns rather than relative duration, although for the other pairs tested, duration does the work. Another study (Bennett, 1968) considered the matter in a comparison of educated northern German speech and the RP accent of British English. The outlook, however, was somewhat different from that of the Swedish research and that of the present report in that Bennett did not fasten his attention upon minimal pairs allegedly distinguished by length. Instead, in both languages, he chose pairs of words differentiated not only by vowel length but also by two degrees of quality; there were short–long pairs near each other in the vowel space, i.e., with small differences in quality, and pairs farther apart in the vowel space, i.e., with greater differences in quality. Anyway, his experiments with synthetic speech yielded a general conclusion of interest here. In both languages, a pair of short and long vowels must be very close in quality for duration to be more important than spectral form (p. 76).

Let us turn now to acoustic and perceptual work on distinctive vowel length in a Southeast Asian language, Thai. For the feature of length in that language, the major acoustic differentiator seems to be relative duration (Abramson, 1962). At the same time, at least one concomitant feature can be detected. In a number of contexts, for some vowel pairs more than others, differences in quality can be heard. That is, the short member is sometimes more open than the long member. This difference in vowel quality, of course, is expected to be reflected in the formant patterns. We have sought to learn whether this other feature bears any of the communicative burden. That is, even if both acoustic features associated with this phonetic distinction are derivable from a single underlying temporal mechanism, it may still be true that listeners, having learned to avail themselves of redundancies in the signal, depend to some extent on more than one acoustic cue. Thus, the question arises as to the relative power of these putative cues. If they both carry phonetic information, are they equal, or is one of them dominant?

In addition, as implied by our remarks on historical linguistics, this work is motivated by a desire to see whether this feature of vowel quality, supposedly concomitant with a primary feature of length, has, as it were, risen in the consciousness of members of the language community to the point of being enhanced in production and established as primary.[1] Indeed, as in the Swedish case mentioned, there might be an asymmetrical condition of a mixture of features in the system of Thai vowels.

Standard Thai, which is based on the Central Thai dialect of the area around Bangkok, has a system of nine vowels that, with some contextual constraints, may be phonologically short or long, as shown here:

i iː ɨ ɨː u uː
e eː ə əː o oː
æ æː a aː ɔ ɔː

Earlier work (Abramson, 1962, 1974) has shown that expressions minimally distinguished in a given context by any of these pairs of vowels are well differentiated physically by relative vowel duration. In addition, experiments with gradually shortened long vowels have demonstrated that relative duration is a *sufficient* cue for perception. There is also the auditory impression of more open quality in the short members of at least some of the pairs; this is supported by formant-frequency differences.

Our hypothesis in this study was that new experiments, designed to allow for the possibility, would show that the vowel spectrum, that is, the formant pattern, which is the major physical correlate of vowel quality, is a significant secondary cue in some of the vowel pairs, especially the pairs /i iː/, /e eː/, /u uː/, and /o oː/. At the same time, relative duration should once again be shown to be the primary cue.

2. ANALYSIS

2.1. Linguistic Material

Five pairs of Thai words, minimally distinguished by vowel length, are shown in Thai script and phonemic transcription with English glosses in the upper portion of Figure 1. These words, representing five of the nine vowel pairs in the Thai system, were embedded in carrier sentences, as seen in the examples at the bottom of the figure, and recorded at normal and fast rates twice by an adult native male speaker of Bangkok Thai.[2] He was a graduate student at the time. The other four vowel-pairs of the system were not included, because we could not find syntactically equivalent minimal pairs for embedding in normal sentences of the vernacular. In the hope of achieving natural speech, for both measurements of production and testing of perception, we wished to avoid rather artificial carrier sentences.) Diphthongs were left out because of the difficulty or impossibility of segmenting them physically into vowels and glides. Of the five tones of the language, only one, the low falling tone, is represented in the figure.

Word-Pairs Recorded for the Tests

จิบ	/cìp/	'to sip'	จีบ	/cìːp/	'to flirt'
เห็ด	/hèt/	'mushroom'	เหตุ	/hèːt/	'cause'
สด	/sòt/	'fresh'	โสด	/sòːt/	'unmarried'
ขุด	/khùt/	'to dig'	ขูด	/khùːt/	'to scrape'
ตัก	/tàk/	'to dip up'	ตาก	/tàːk/	'to dry'

Samples of Words in Carrier Sentences

อย่าขุดมากเกินไป /jàː khùt mâːk kəːn paj/
'Don't dig too much.'

อย่าขูดมากเกินไป /jàː khùːt mâːk kəːn paj/
'Don't scrape too much.'

ไม่ทราบสดหรือเปล่า /mâj sâːp sòt rɨ̂ plàːw/
'I don't know whether it's fresh.'

ไม่ทราบโสดหรือเปล่า /mâj sâːp sòːt rɨ̂ plàːw/
'I don't know whether he's single.'

Figure 1. Pairs of Thai words, given in Thai script and phonemic transcriptions with glosses, used in the experiments. Two carrier sentences with four of the words embedded are also shown.

This was because we wanted to limit our syllable types, for ease of handling in the perceptual work, to those with final stop consonants.[3] Such syllables do not permit the occurrence of all five tones. In fact, for short and long vowels in this context, the Thai lexicon is best represented by words on the low falling tone. The speaker, as well as another Thai person and the first author, who has had long practical experience with the language, listened to the recordings and found them acceptable as fluent utterances of ordinary colloquial speech.

The durations of the vowels of the key-words, from the release of the first consonant to the closure of the second, were measured from waveforms. A waveform-editing computer program enabled us to magnify waveforms or any portion thereof for greater precision in marking these points. For the pair beginning in /h/, of course, there was no oral closure or constriction at the onset, so, taking the phonetic manifestation of /h/ to be simply noise-excitation of the first portion of the vowel, we marked the beginning of the [h]-turbulence as the beginning of the vowel. FFT analyses were applied to the midpoint of each vowel to measure the frequencies of the first three formants. In addition, narrow phonetic transcriptions, with reference to the IPA Cardinal Vowels, were made by the first author in two sessions separated by three weeks.

2.2. Results

In Table 1 are given the durations and formant frequencies of the vowels of the male speaker uttered in the carrier sentences at his normal conversational rate. His long vowels were on average 1.9 times longer than the short vowels. This is much as in earlier data (Abramson, 1962; Gandour, 1984; Gandour, Weinberg, Petty, & Dardarananda, 1987; Sittachit, 1972). Within each length category, the two available tokens differ somewhat in duration, in conformity with data from larger samples (e.g., Abramson, 1962); nevertheless, there is no overlap between short and long vowels. For the front vowels in the table, the short vowels generally have higher F_1 values and lower F_2 values than their long counterparts. For the back vowels, not only is F_1 higher in the short members but so is F_2. For the pair /a aː/, on the other hand, no trend is apparent in the first two formants; all that can be said is that F_3 is higher in the long vowel.

Table 1. *Durations and formant frequencies, each rounded to the nearest 5 ms and 5 Hz respectively, for two tokens of each vowel recorded by a male speaker for the perception experiments. Asterisks mark the tokens actually used in the tests.*

Word	Vowel Duration (ms)	F_1	F_2 (Hz)	F_3
/cìp/$_1$	30	315	2025	2740
/cìp/$_2$*	35	320	2030	2530
/cìːp/$_1$	140	250	2185	2895
/cìːp/$_2$*	105	295	2115	2745
/hèt/$_1$	135	390	1965	2500
/hèt/$_2$*	130	440	1975	2555
/hèːt/$_1$*	195	390	2035	2565
/hèːt/$_2$	285	370	2100	2565
/tàk/$_1$	100	880	1505	2580
/tàk/$_2$*	85	810	1580	2660
/tàːk/$_1$	135	825	1560	2705
/tàːk/$_2$*	150	850	1565	2715
/khɯ̀t/$_1$*	125	350	1110	2370
/khɯ̀t/$_2$	90	385	1100	2490
/khɯ̀ːt/$_1$*	215	300	750	2570
/khɯ̀ːt/$_2$	200	310	850	2490
/sɔ̀t/$_1$	65	490	1145	2665
/sɔ̀t/$_2$*	55	470	1165	2550
/sɔ̀ːt/$_1$	90	465	1050	2555
/sɔ̀ːt/$_2$*	125	455	880	2635
Av. V	85			
SD	36			
			Vː/V=1.9	
Av. Vː	165			
SD	56			

The auditory phonetic assessment yielded an impression of slightly more open qualities for the short members of the pairs, together with slightly more rounding for long /ɯː/ and /ɔː/ than for their short counterparts. For the pair /a aː/, however, there was no consistent difference in transcription.

3. PERCEPTION

3.1. Experiments

As a baseline for the experiments in perception, control tests for the identification of the original words in the carrier sentences spoken at both rates were prepared. That is, all the items in

Table 1 together with the fast versions mentioned in 2.1 were recorded directly from the computer files onto magnetic tape to make two randomized test orders for every word pair, with three tokens of each stimulus in each test. This means that, item by item, the subjects never knew whether to expect a short or a long vowel and whether the key word would be embedded in a carrier sentence uttered at a normal or a fast rate. The resulting five tests, in two test orders each, were played through headphones to 50 undergraduates at Chulalongkorn University in Bangkok. All of them had been screened by the staff of the Department of Linguistics as native speakers of Bangkok Thai. The students, giving a total of six responses to each original production, identified the key words in Thai script.

The rest of the listening tests made use of the sentences recorded at normal speed. The utterances that were altered to make stimuli are marked with asterisks in Table 1. In the circumstances, with a choice of only two tokens of each word, we considered the durations of the long and short members of each pair, their formant patterns, and the impressionistic transcriptions to choose a pair most likely to have audible differences in both length and quality for our listeners. Favoring one combination over another was a fairly subtle matter and for one pair, /tàk/ vs. /tàːk/, perhaps arbitrary.

To make our stimuli, we lengthened the short member of each of the five pairs in steps of about 10 ms up to a duration a bit longer than that of its long counterpart. Likewise, we shortened the long member in steps of about 10 ms down to a duration a bit shorter than that of its original short counterpart. In order to avoid interfering with any formant movements present, we used spectrograms to help find a point early in the steady-state voice-excited portions of the vowel formants at which to begin either procedure. To lengthen an original short vowel, we copied one period of the waveform at this point and then repeatedly inserted it at the end of the one copied, only at zero-crossings, thus making longer and longer variants. To shorten an original long vowel, we began at the same point and removed one period after another after the first one, cutting only at zero-crossings, thus making shorter and shorter variants. No distortions detectable either by ear in the signal or by eye in the waveform were introduced. Indeed, even the fundamental-frequency pattern was not disrupted, since this early in the syllable the slight downward contour of the low falling tone—called simply "low" by others—had not yet begun or had barely begun. That is, all the stimuli sounded like natural human speech, although we were sure that a number of them in each series would be ambiguous for Thai listeners as to phonemic length category. All the variants were kept within the carrier sentence of the original word from which they were derived.

With these methods, we were able to pit possible spectral cues against durational cues. That is, we kept the original formant patterns of the vowels while altering the durations. Our hypothesis was that a lengthened short vowel should have a later crossover point in identification functions than its shortened counterpart. We supposed that spectral features of each length category would expand the range of that category over durational variants in the region that would otherwise be ambiguous. Thus, the category boundary should be pushed in the opposite direction of the starting category by spectral cues.

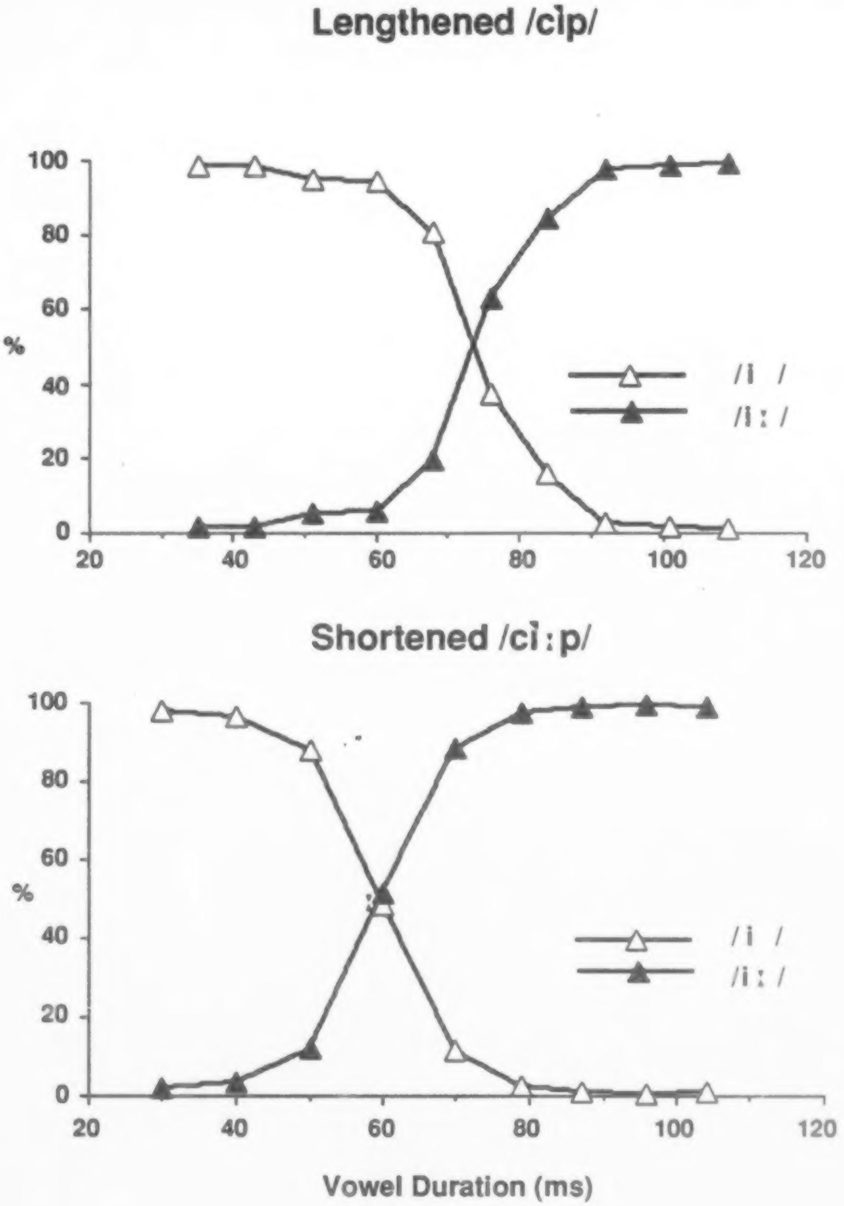
A listening test was prepared in three randomizations for the set of alterations of each original word together with the original word itself. A sampling of the range of stimuli was presented before each test. Each of the three test orders contained three tokens of each stimulus embedded in its carrier sentence. Written and oral instructions were given in Thai. Our 50 subjects listened through headphones and labeled each stimulus a total of nine times as one of two possible words given to them in Thai script on the answer sheets. The subjects were split into three groups for counterbalancing in the order of taking combinations of these tests and the control tests. Each group took the tests in four sessions of less than one hour each. Imperfect attendance at some of the sessions is reflected in the data reported in 3.2.

3.2. Results

We shall not display the results of the control tests, because, with a few scattered confusions at the fast rate, the subjects gave virtually 100% correct identification of all the stimuli. That is, the vowel-length distinction is still robust in Standard Thai, at least for the contexts under consideration. This means also that all the words chosen as bases for the making of stimuli were themselves completely unambiguous as to phonemic length.

The perceptual results of the experimental manipulations of the short and long vowels are given in Figures 2–6. The two coded lines in each graph give us the percentage identifications of the stimuli as either short or long vowels. Since our procedures did not yield the same number of stimuli across the two conditions shown in each figure, i.e., lengthening one member and shortening the other, we extracted the 50% crossover points from cubic polynomial curves fitted to the curves for the labels as short vowels.

In Figure 2 are seen the results for the experiments with the vowels /i iː/. While in both of these tests duration is obviously a sufficient cue, the average perceptual crossover points are significantly different : t (46) = 17.3, $p < 0.001$. For lengthened /cìp/, the crossover is at 74 ms and for shortened /cìːp/, 59 ms. That is, the zone of ambiguity is later for the original short vowel and earlier for the original long vowel.



Lengthened /cìp/

Shortened /cìːp/

Figure 2. Percentages of labels as short or long vowels of 47 subjects for variants of lengthened / i / and shortened / iː /.

In Figure 3, the crossover point for lengthened /hèt/ is at 167 ms; for shortened /hèːt/ it is at 140 ms. Here again, we see the same shift: t (48) = 24.9, $p < 0.001$.

In Figure 4, the crossover point for lengthened /sòt/ is at 85 ms; for shortened /sòːt/ it is at 79 ms: t (48) = 4.9, $p < 0.001$.

For the one pair in which the vowels seem so much the same to the ear that no difference in crossover was expected, we get the same kind of results, as seen in Figure 5. For lengthened /tàk/ the crossover point is at 108 ms; for shortened /tàːk/ it is at 104 ms: t (48) = 6.2, $p < 0.001$.

Examination of the data in Figure 6 shows that for lengthened /khùt/, the crossover point is at 166 ms; for shortened /khùːt/ it is at 175 ms: t (48) = 8.7, $p < 0.001$. Although the difference is significant, as in all these experiments, the direction is reversed! That is, the crossover point for the lengthened short vowel is earlier than that of the shortened long vowel.

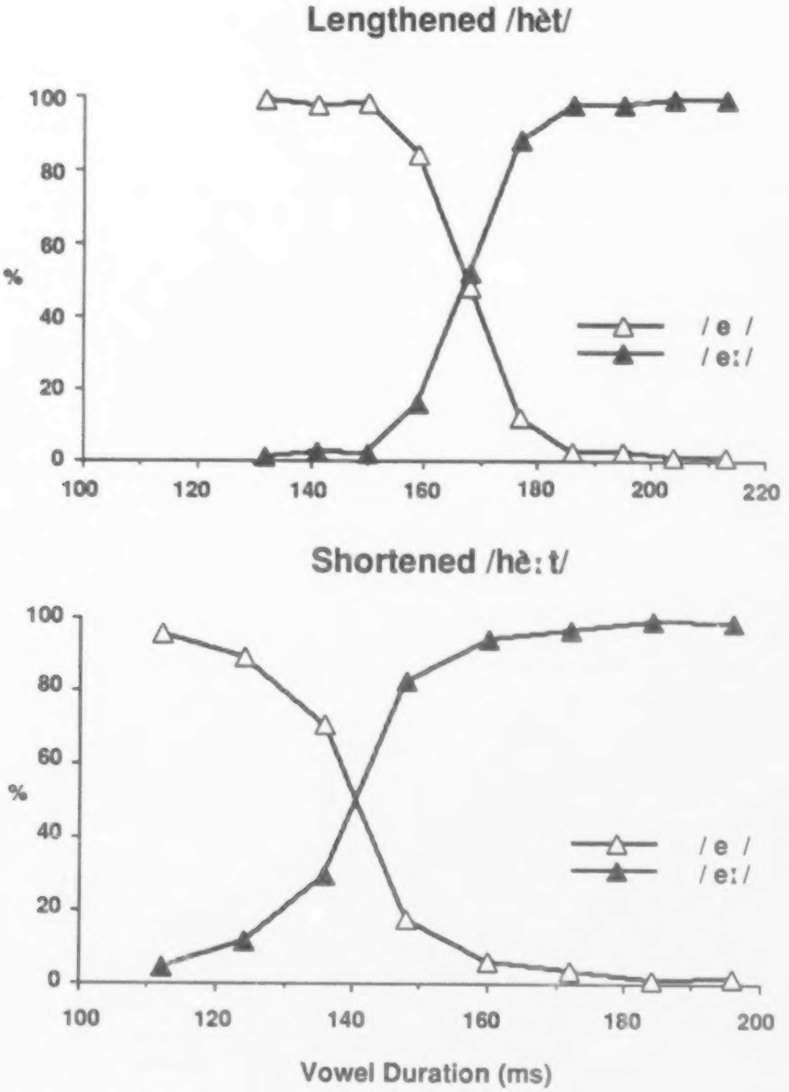
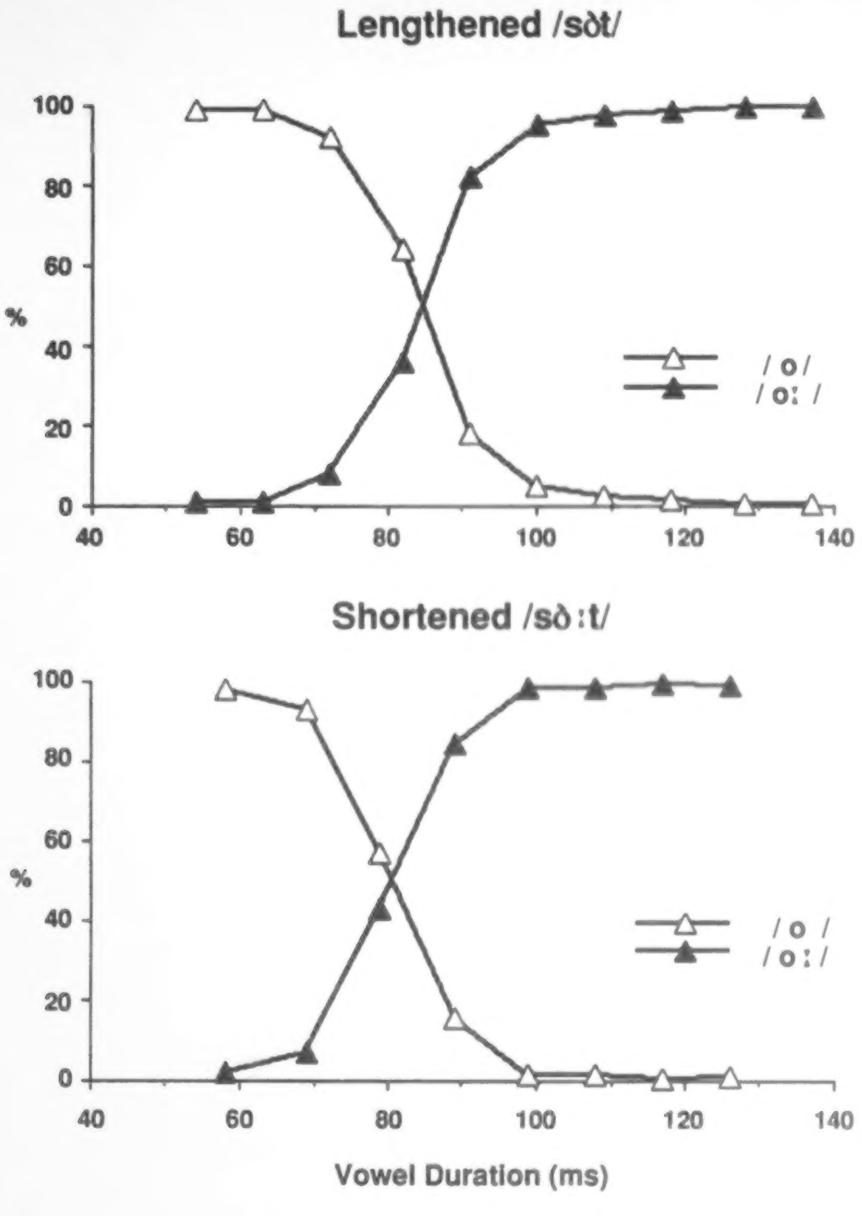


Figure 3. Percentages of labels as short or long vowels of 49 subjects for lengthened / e / and shortened / eː /.



Lengthened /sɔ̀t/

Shortened /sɔ̀ːt/

Figure 4. Percentages of labels as short or long vowels of 49 subjects for lengthened / e / and shortened / eː /.

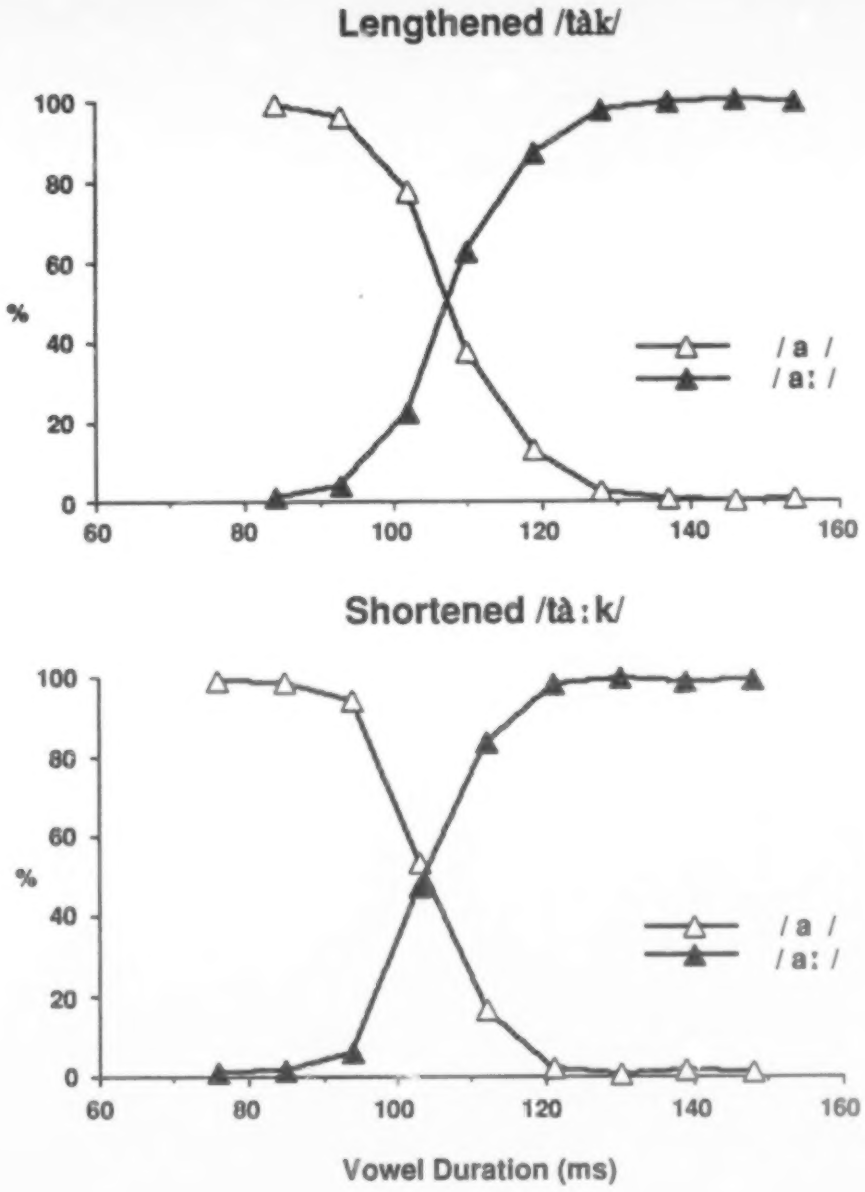


Figure 5. Percentages of labels as short or long vowels of 49 subjects for lengthened / a / and shortened / aː /.

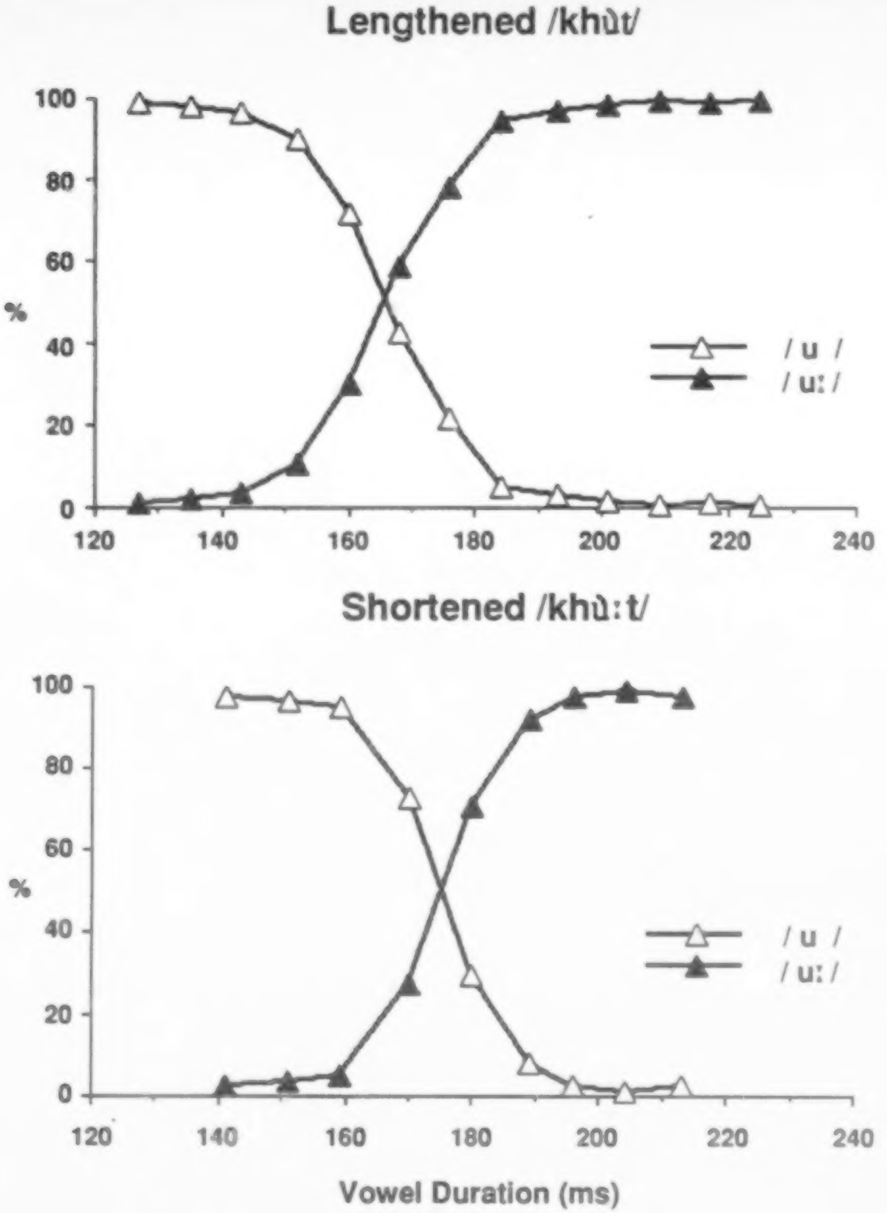


Figure 6. **Percentages of labels as short or long vowels of 49 subjects for lengthened / u / and shortened / uː /.**

4. DISCUSSION AND CONCLUSION

Nearly 30 years after the first experiments (Abramson, 1962), it is evident that relative duration is still the dominant cue to distinctive vowel length in Thai. This, of course, will not surprise linguists who have seen no reason to doubt the traditional view that quantity is a feature of the Thai vowel system. The effects of speaking rate on this distinction have been investigated by Katyanee Svastikula (1986).

For all the pairs tested here, and probably for the whole vowel space, features other than length provide less powerful but pertinent cues. Our hypothesis was that a major secondary cue to the distinction lies in spectral differences between the long and short members of the vowel pairs. Even on this hypothesis, incrementally lengthening an original short vowel would at some point cause listeners to start hearing its long counterpart; nevertheless, the perceptual crossover zone ought to be delayed to the extent that spectral cues to "shortness" persisted. Likewise, incrementally shortening an original long vowel should induce a percept of shortness only after the spectral features of the long vowel have been overcome by sufficient decreases in duration; thus, the crossover point will be earlier in the syllable than for the lengthened short vowel. That is, in both conditions, the shift between categories will be determined mainly by relative duration, but the exact placement of the boundary will be affected by secondary cues in the spectrum. This hypothesis is well supported by four out of the five pairs tested. In all of them, duration is a sufficient and powerful cue, but the placement of the boundary is significantly later for the lengthened short vowels. The pair /u uː/, however, is a puzzling exception. It is *earlier* for the series of stimuli made by lengthening the short vowel. Although this difference is statistically significant too, the direction is reversed. Once we assured ourselves that there had been no error in our records or in the manipulation of these items, we had to seek other causes. One possibility is rhythmic factors in the carrier sentences. Given a sentence that can syntactically and semantically contain either member of a pair of words minimally distinguished by the feature of vowel length, perhaps there is a tendency toward reciprocity between the duration of the vowel in the key word and the duration of the whole sentence. That is, the sentences containing short vowels will be longer than their counterparts containing long vowels. We have evidence in support of this in Table 2. On the left, we see the average perceptual crossover point for the stimuli derived from each vowel and the difference between the crossovers for each word pair. On the right, we see the durations of the carrier sentences *without* the key words and the differences between them for each pair. For the first four pairs of words, as predicted and already described, the crossover points are earlier for the original short vowels, and the carrier frames are slightly longer. For the pair in question, however, /u uː/, it is just the opposite. It must be admitted that the carrier-duration differences for the first four pairs are so small that it is psychoacoustically unlikely that they have much, if any, perceptual effect, although they are consistent with the frames of the unused items in Table 1. The negative number for the carriers of the last pair, however, is relatively large. Inasmuch as it would require further experimentation to assert that this is the relevant factor for explaining the discrepancy, we can say at this time only that it is suggestive.[4]

Table 2. *Perceptual crossover points for the pairs of source words, with differences; durations of the carrier sentences without the key words, with differences. V = short vowel; V: = long vowel. All values are in ms.*

	PERCEPTION		CARRIER SENTENCE	
Source	Crossover	V – V:	Duration	V – V:
cìp	74		1046	
cìːp	59	15	1037	9
hèt	167		1454	
hèːt	140	27	1449	6
sòt	85		1177	
sòːt	79	6	1168	7
tàk	108		983	
tàːk	104	4	976	7
khùt	166		1300	
khùːt	175	–9	1333	–33

Since relative duration is the dominant cue, it seems likely that the underlying mechanism is timing of the vowel articulation, with concomitant small differences in vocal-tract shape. The latter give rise to sufficient differences in formant pattern to yield slight but audible differences in vowel quality in association with the length feature. Thus, we seem to have here one more case of a combination of acoustic cues, differing, to be sure, in relative power, underlying a single distinctive feature. A troubling question is why a vowel-length distinction should entail any differences in the dimensions of the resonating cavities. After all, in the Introduction we pointed out that one ought to be able to hold a particular setting of the articulators for a shorter or longer time, presumably without any change in position. Here again, stimulated by observations in the literature (e.g., Durand, 1946; Lehiste, 1970, pp. 30–33), we can only speculate. If in the history of a language, a phonological distinction ostensibly based on a single phonetic feature comes into being, we generally believe that the contrasting phonemes are entered into the mental grammar with a special bond between them. That is, there is a kind of psycholinguistic reality to their status of being differentiated by a single feature rather than more than one. Thus, for example, English /b/ and /p/ would be more closely linked than either of them with /f/. Another possibility, however, is that each phoneme takes on a psychological life of its own that allows it, under mostly unexplained pressures, to accrue to itself certain bits of phonetic behavior not obviously attributable to its setting on the allegedly "distinctive" dimension. Perhaps these accrued characteristics result from not fully understood forces, tensions, and inertias of the vocal apparatus. In the particular case of distinctive vowel length, then, we might imagine that over a long period of transmitting a quantity language from generation to generation, speakers may come to produce steady-state long vowels with articulatory settings slightly different from those of the short counterparts, giving each member of a minimal pair something of a phonetic life of its own. This may happen faster in some parts of the vowel system than others. If this is happening, perhaps through natural forces governing speech events, these small differences will be audible and can provide redundant information in support of what may still be the old primary cue.

In such situations, we are probably observing a potential for a diachronic shift from length to quality (see, e.g., Hadding-Koch & Abramson, 1964) in at least part of the system and, maybe ultimately, in the whole system. Indeed, our understanding of the historic rise and fall of distinctive vowel length (de Chene, 1985) may be helped by further experimental investigations.

Finally, for the four pairs of vowels in which the formant pattern seems to provide secondary cues to the distinction, it has to be noted that the magnitudes of the spectral differences have no bearing on the level of significance of the differing crossover points for the various pairs of vowels. On the other hand, the absolute values of the differences do differ somewhat across the pairs, for example the pair /a aː/ with its small spectral difference but its unexpectedly significant difference in crossover points, but it is hard to say much about this without a better metric for quality judgments and their correlation with formants variously distributed in auditory space.

An approach yet to be taken is to introduce spectral variants into each durational variant in our experiments. That is, for each of the stimuli used here, we ought to change the formant frequencies in small steps for further perceptual testing. Our preliminary attempt at doing this with an older computer program yielded promising tentative results with Thai subjects but also caused disturbing enough artifacts for us to discard these tests. In an extension of the present research, we plan to use a newer technique of this sort in the hope of shedding further light on the topic.

REFERENCES

Abramson, A. S. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments. (Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20 [*International Journal of American Linguistics, 28*, No. 2, Part III]). Bloomington.

Abramson, A. S. (1974). Experimental phonetics in phonology: Vowel duration in Thai. *Pasaa: Notes and News about Language Teaching and Linguistics in Thailand, 4*, 71–90.

Abramson, A. S. (1986). The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association, 16*, 8–16.

Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. In *Proceedings of the XIth International Congress of Phonetic Sciences, 6* (pp. 66–70). Tallinn, Estonia: Academy of Sciences of the Estonian S.S.R.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech, 11*, 65–85.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.

Catford, J. C. (1988). *A practical introduction to phonetics*. Oxford: Oxford University Press.

de Chene, B.E. (1985). *The historical phonology of vowel length*. New York: Garland.

Durand, M. (1946). *Voyelles longues et voyelles brèves: Essai sur la nature de la quantité vocalique*. Paris: Librairie C. Klincksieck.

Gandour, J. (1984). Vowel duration in Thai. *Crossroads, 2*, 59–64.

Gandour, J., Weinberg, B., Petty, S. H., & Dardarananda, R. (1987). Vowel length in Thai alaryngeal speech. *Folia Phoniatrica, 39*, 117–121.

Garnes, S. (1973). Phonetic evidence supporting a phonological analysis. *Journal of Phonetics, 2*, 273–283.

Hadding-Koch, K., & Abramson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica, 18,* 94–107.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Sittachit, K. (1972). Observations on short and long vowels in Thai. In J. G. Harris & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 28–32). Bangkok: Central Institute of English Language (in Thai).

Straka, G. (1959). Durée et timbre vocalique. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft, 12,* 276–300.

Svastikula, K. (1986). *A perceptual and acoustic study of speech rate in distinctive vowel length in Thai*. Unpublished doctoral dissertation, The University of Connecticut, Storrs.

FOOTNOTES

[*] *Journal of Phonetics, 18(2),* 79-92 (1990).

[†] Also University of Connecticut Department of Linguistics, Storrs.

[1] Another language under study, Pattani Malay, raises similar questions for its word-initial distinctive consonant length (Abramson 1986, 1987).

[2] Three Thai women were recorded in the same way, but their utterances were not used here.

[3] The only other available endings are nasal consonants,vowels, and glides. These were rejected in this work as awkward for later experimental manipulation.

[4] One obvious extension of this work is to run the same experiment with the other token of /khùa/ in Table 1. Its carrier frame has a duration of 1030 ms.

Haskins Laboratories Status Report on Speech Research
1990, SR-101/102, 269-287

Effects of Preceding Context on the Voice-Onset-Time Category Boundary*

Bruno H. Repp and Hwei-Bing Lin†

The category boundary on a voice-onset-time (VOT) continuum ranging from *bin* to *pin* shifts when the stimuli are preceded by different carrier phrases. By employing a variety of precursor phrases, by varying the temporal interval between precursor and test word, and by selectively eliminating either voicing information or spectral structure from the precursors, the present experiments show that the context effect is caused by the presence or absence of voicing in the precursor's final segment. The effect decreases with temporal separation but persists over several seconds. In addition, the "baseline" VOT boundary for *isolated* test words interspersed in a test sequence shifts depending on what precursor stimuli occur in the same sequence. The perception of VOT thus seems to be sensitive to both close and distant manifestations of laryngeal activity, always in an assimilative fashion, which suggests an integrative perceptual mechanism with a long time constant.

INTRODUCTION: AN UNEXPECTED BOUNDARY SHIFT

The research described in this paper was stimulated by an unexpected finding. In recent experiments (Repp & Lin, 1989) we investigated the effect of preceding context on the discriminability of small differences in voice onset time (VOT). The subjects in one of our experiments (Experiment 3) also were given an identification test in which stimuli from a *bin-pin* continuum varying in VOT were presented either in isolation or preceded by one of two carrier phrases, *Take this* or *Take the*. (These stimuli will be described in more detail below.) The results of this test, which were not reported previously, are shown in Figure 1. There was a clear shift of the category boundary in favor of *pin* responses following *Take this*, whereas the boundary following *Take the* was in the same place as that for isolated test stimuli. The average category boundaries, estimated by probit analysis and averaged over subjects, were at VOT values of 35.2 ms (no precursor), 26.6 ms (*Take this*), and 36.7 ms (*Take the*). The difference among these three boundaries was highly significant [F(2,24) = 83.99, $p < .0001$], but there was no significant boundary difference between the no-precursor and *Take the* conditions compared separately.

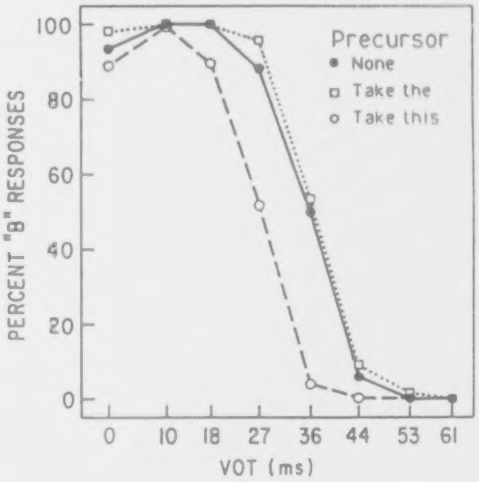


Figure 1. Percentage of "B" responses as a function of VOT for stimuli from a *bin-pin* continuum in three precursor conditions. Average data of thirteen subjects.

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories.

This type of context effect has, to the best of our knowledge, not been described previously in the literature. We decided to investigate its cause, as this was likely to tell us something new and important about the perception of the phonetic voicing distinction, a topic of long-standing interest in speech research (see, e.g., Lisker & Abramson, 1970; Pisoni, 1977; Summerfield, 1981; Berg, 1986). Phonetic context effects are of theoretical significance because they reveal either listeners' tacit knowledge of systematic contextual variation in speech production (cf. Repp, 1982) or interactions in the auditory processing of the speech signal (cf. Diehl & Kluender, 1989), both of which are essential ingredients of any model of speech perception.

Most experiments on the perception of the voicing distinction in syllable-initial stop consonants have used isolated syllables or words, so the critical segment was in fact utterance-initial. A few studies, however, have demonstrated effects of preceding context. The most relevant of these was conducted by Summerfield (1981) who showed that the simulated speaking rate of a precursor phrase affects the perceptual boundary on a VOT continuum, slower precursor rates leading to an increase of responses in the voiced category. Summerfield also was able to show that this effect derived primarily from the temporal variations in the final syllable of the precursor (however, see Kidd, 1989), and that it decreased substantially as the silent interval between precursor and test syllable was lengthened beyond a duration acceptable for a stop consonant closure. The duration of that closure itself had a rate-dependent effect on the perceived stop voicing, relatively shorter closures leading to more voiced responses (again, see Kidd, 1989, for different results). Also, the precursor effects were unidirectional when compared to a no-precursor baseline: The VOT boundary shifted towards longer values when the precursor rate was slow, but it stayed in place when the precursor rate was fast. The precursor phrase ended with the word *the* in these experiments. Summerfield's findings are thus consistent with the absence of a boundary shift in our (moderately fast) *Take the* precursor condition relative to the no-precursor baseline (Figure 1), but they cannot explain the shift in our *Take this* condition, where the precursor was produced at the same intended rate.

This conclusion may be premature, however. The precursors in our experiment had been excerpted from complete phrases of *Take this bin* and *Take the pin*, and they included the original silent closure intervals. Even though the boundary difference obtained in the perceptual test was contrary to what might be predicted if there had been coarticulatory information predictive of stop consonant voicing, it is worth noting that the closure durations differed between the *Take this* (115 ms) and *Take the* (84 ms) precursors. Since longer closures are generally associated with voiceless stop consonants (e.g., Lisker, 1957), this difference could easily account for the boundary shift in favor of *pin* responses following *Take this*, even though that phrase was originally followed by *bin*. In fact, Summerfield (1981) found such an effect of closure duration in his Experiment 3. Our first order of business, therefore, was to rule out this relatively trivial explanation of our finding.

EXPERIMENT 1: THE ROLE OF CLOSURE DURATION

In Experiment 1 we compared the effects of the two precursors at equal closure durations. To study the perceptual contribution of closure duration itself, we used four different silent intervals that spanned the normal range of stop consonant closures in this context.

In addition, we included a third precursor, referred to as *Take this-*. Its final /s/ noise had originally been produced in word-initial position (viz., in the word *spin*), but had been excerpted and substituted for the original /s/ noise of *Take this*. Since English phonology permits only /p/ following a word-initial /s/, it was of interest to see whether subjects' labeling responses would be shifted even more towards *pin* in the *Take this-* context, particularly at the shorter closures which characterize fricative-stop clusters (cf. Wingate, 1982). When we used the same stimuli in a VOT discrimination task (Repp & Lin, 1989), we found much more interference from the *Take this-* precursor than from the *Take this* precursor. This made us suspect that the location of the word boundary was ambiguous in the former context, leading to occasional perception of *Take the spin*. It is conceivable that, at short closure durations, even *Take this* with a proper word-final /s/ permits some word boundary ambiguity; if so, this would provide an alternative explanation of the boundary shift observed in our pilot study (Figure 1). Such an effect should disappear at longer temporal separations, however.

The predictions for Experiment 1, then, were as follows: If the previously observed VOT boundary difference following *Take this* and *Take the* was entirely due to the difference in closure durations, it should be absent at all closure durations. If it was due to word boundary ambiguity in the *Take this* context, it should be present at short closures but disappear at longer ones. This tendency should be even stronger in the case of the *Take this-* precursor. In addition, if closure duration provides a secondary cue to stop voicing, *pin* responses should increase with closure duration, up to a point.

Methods

Subjects

Thirteen paid volunteer subjects from the Yale community participated. They were all native speakers of American English and reported having normal hearing.

Stimuli

The *bin-pin* continuum was constructed from natural productions of these words in the context of *Take this* by a female speaker. The utterances were recorded in a sound-insulated booth, digitized with a 20 kHz sampling frequency, low-pass filtered at 9.6 kHz, and edited to separate the final words from the carrier phrases. Using a waveform editor, a VOT continuum was created by replacing initial periodic segments of *bin* with aperiodic segments of equal duration from *pin*, proceeding in steps of two glottal cycles. The six stimuli used in Experiment 1 had VOTs of 10, 18, 27, 36, 44, and 53 ms (rounded to the nearest ms). These stimuli were the same as in our pilot study (Figure 1), except that the two extreme stimuli of the earlier continuum were omitted for reasons of economy.

The precursors, too, were the ones used previously, spoken by the same person. The phrase *Take this* had originally been produced before *bin*, while *Take the*, to avoid continuation of voicing into the closure, had been excerpted from an utterance ending in *pin*. (Possible effects of anticipatory coarticulation will be considered later; we have noted already that the direction of the observed perceptual boundary shift was contrary to what might be predicted on that basis.) The *Take this-* precursor, previously used in Experiment 2 of Repp and Lin (1989), was acoustically identical with *Take this* except for the spliced-in final fricative noise which had been excerpted from the utterance *A crazy spin*. That word-initial /s/ noise was longer than the word-final one it replaced (148 ms versus 94 ms), and probably also had spectral and amplitude characteristics that reflected its provenance.

The duration of the silent closure interval following each precursor was varied in four steps: 45, 80, 115, and 150 ms. Each of the six *bin/pin* stimuli was presented in isolation and preceded by each of the three precursors at each of the four closure durations, which resulted in 78 stimuli that were recorded five times in different random orders, with interstimulus intervals of 2 s.

Procedure

The subjects listened binaurally in a quiet room over TDH-39 earphones at a comfortable intensity. The task was to identify the initial consonant of each test word as "B" or "P."

Results and Discussion

The average labeling functions (percent "B" responses as a function of VOT) are shown separately for each closure duration in the four left-hand panels of Figure 2. It is evident, first, that the *Take the* precursor condition yielded results very similar to those for isolated stimuli, which replicates our earlier findings. Second, there was a large boundary shift in the *Take this* precursor condition, favoring voiceless stop percepts at all closure durations. The shift was even much larger than in the pilot experiment (Figure 1). Since the average boundary location in the *Take this* condition (29 ms of VOT) was similar to that obtained previously (27 ms of VOT), the larger boundary shift resulted from an unexpected increase in voiced stop responses in the *Take the* and no-precursor conditions.[1] A third finding is that, as predicted, the *Take this-* precursor introduced an additional bias to report voiceless stop consonants at short VOTs. Finally, it is obvious that variations in closure duration had surprisingly little effect.

An analysis of variance, with precursor condition and closure duration as factors, was conducted on subjects' proportions of *bin* responses (averaged over VOT) in the three precursor conditions. It revealed highly significant effects of condition [F(2,24) = 142.84, $p < .0001$, MSe = 26.30] and of closure duration [F(3,36) = 8.79, $p < .0003$, MSe = 3.25], as well as an interaction between the two [F(6,72) = 7.33, $p < .0001$, MSe = 2.52]. The latter two effects reflect the fact that there was an increase in voiced responses with increasing closure duration, but only in the *Take this* and *Take this-* conditions. This resulted in a reduction of the boundary shift with increasing closure duration, though it was still very large at the longest interval used.

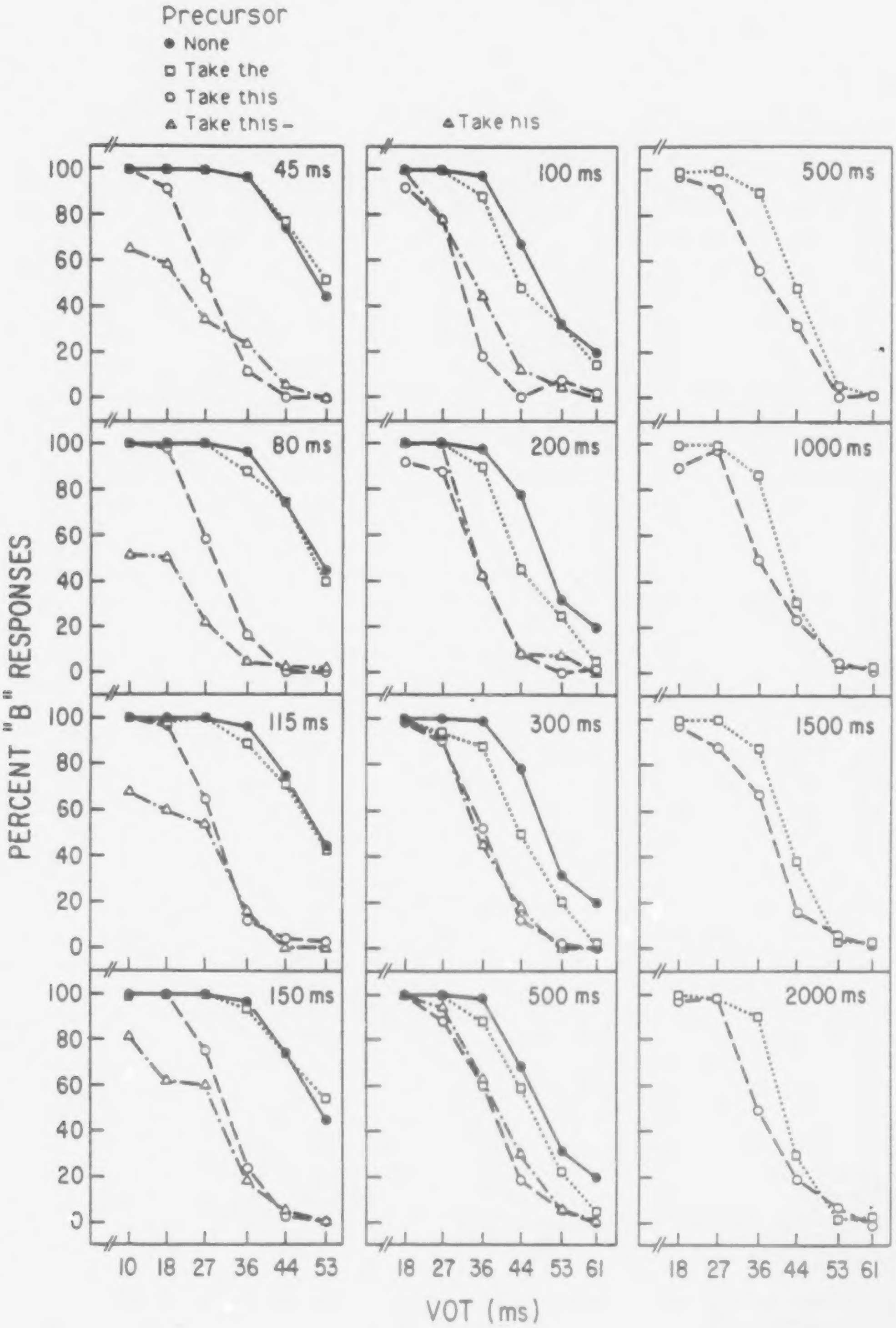


Figure 2. **Results of Experiment 1 (left-hand panels), Experiment 2 (center panels), and Experiment 3 (right-hand panels). Each panel represents a particular closure duration.**

These results definitely eliminate the hypothesis that our original observation of a boundary shift was caused by differences in silent closure duration associated with the precursors. In fact, closure duration had an unexpectedly small effect. In the *Take the* context it did not affect the subjects' responses at all, contrary to Summerfield's (1981) finding of a decrease in voiced responses at intervals of 100-250 ms in a comparable situation. In the *Take this* and *Take this-* contexts, there was, on the contrary, an increase in voiced responses with closure duration. This should probably not be interpreted as a direct effect of closure duration as a cue to stop consonant voicing (cf. Kidd, 1989), but rather as a reduction with increasing temporal separation of some effect exerted by the precursors on the test syllables.

What, then, was responsible for this effect? One possibility was considered earlier: ambiguity of the word boundary location. It was expected, however, that this factor would come into play only at relatively short closure durations, those characteristic of tautosyllabic fricative-stop clusters (less than 100 ms, say). The finding that the boundary shift was still so large at the longest closure duration (150 ms) seems to argue against this explanation. However, the additional effect exerted by phonetic ambiguity in the *Take this-* precursor, obtained as expected at short closures, also persisted at the longest closure in only slightly attenuated form. Thus, a role of word boundary ambiguity cannot be ruled out completely. In any case, the difference between the *Take this* and *Take this-* conditions demonstrates that phonetic properties of the precursor's final segment can affect stop consonant perception over separations of at least 150 ms. Perhaps, then, the difference between the *Take this* and *Take the* conditions is also to be attributed to properties of the segments immediately preceding the closure—e.g., to the fact that /s/ is voiceless while /ə/ is voiced. Before examining this hypothesis in detail, we conducted two additional experiments on the effects of temporal separation.

EXPERIMENT 2: LONGER CLOSURE DURATIONS

The results of Experiment 1 raise the question of how long the temporal separation could be before the precursor effect disappears. This question has some bearing on the explanation of the effect: Word boundary ambiguity becomes an unlikely explanation once the closure interval is so long as to become effectively a pause separating the words. Experiment 2 increased the temporal separations up to 500 ms, which far exceeds the range of stop closure durations encountered in natural speech.

In addition, the experiment included a new precursor, *Take his*, which replaced *Take this-*. It was intended to provide preliminary information about whether it was the voicelessness of the final fricative in *Take this* that caused the boundary shift. If so, *Take his*, which ends with the voiced fricative /z/, should not cause any boundary shift.

Methods

Subjects

Eight subjects from the same subject pool participated. Four of them had also served in Experiment 1.

Stimuli

The stimuli were the same as in Experiment 1, with the following modifications. To better accommodate the category boundaries in all conditions, the range of VOTs was shifted by omitting the 10-ms VOT stimulus and adding a 61-ms VOT stimulus. Four longer closure durations were used: 100, 200, 300, and 500 ms. Instead of the *Take this-* precursor, a newly recorded precursor, *Take his*, was used, produced by the same female speaker, originally preceding *bin*. Number of stimuli, test arrangement, and procedure were the same as in Experiment 1.

Results and Discussion

The labeling functions are shown in the center panel of Figure 2, separately for each closure duration. The shift in the range of VOTs apparently had no effect on the no-precursor labeling function. In contrast to Experiment 1, the *Take the* precursor led to a slight reduction in "B" responses relative to the isolated test words. As before, however, the *Take this* precursor caused a much larger boundary shift to the left which diminished somewhat with increasing silent interval duration but, astonishingly, was still clearly present at the 500 ms interval. Finally, contrary to the predictions, the effect of the *Take his* precursor was very similar to that of the *Take this* precursor.

The statistical analysis of the precursor data confirmed the main effect of precursor condition [$F(2,14) = 16.27$, $p < .0003$, $MSe = 22.79$] and its interaction (i.e., its decrease) with closure duration [$F(6,42) = 3.79$, $p < .005$, $MSe = 2.50$]. There was also a main effect of closure duration [$F(3,21) = 5.95$, $p < .005$, $MSe = 3.38$], which reflects the fact that "B" responses increased overall due to the progressive reduction of the *Take this* and *Take his* precursor effects. A separate comparison of the *Take this* and *Take his* conditions showed no difference between them, nor any interaction with closure duration, only a main effect of closure duration [$F(3,21) = 11.52$, $p < .0002$, $MSe = 2.89$], due to a systematic increase in "B" responses.

This experiment yielded two surprising findings. First, the boundary shift was still present at a temporal separation of 500 ms. This rules out an explanation in terms of word boundary ambiguity, though the possibility remains that such ambiguity adds to the effect at short closure durations. Second, the *Take his* precursor had the same effect as the *Take this* precursor, which seems to suggest that the boundary shift was not caused by the voicelessness of the final fricative in *Take this*. However, a closer look at the waveform of the *Take his* precursor clarified this result. It emerged that its final fricative noise, 77 ms in duration, was largely voiceless, even though it had preceded *bin* in the original utterance: Glottal pulses ceased about 20 ms into the noise. Such final fricative devoicing at a juncture preceding another voiced segment appears to be not uncommon, though it is not well documented with data in the phonetic literature, which is unanimous in stating that devoicing occurs when the following segment is voiceless. (For data on British English, see Haggard, 1978.) Our results thus show only that the *phonological* voicing status of the final fricative (/z/ or /s/) is irrelevant to the precursor effect. The hypothesis that the precursor effect derives from the *phonetic* voicing status of the final segment is alive and well but remains to be tested properly (see Experiment 4 below).

EXPERIMENT 3: VERY LONG CLOSURE DURATIONS

The unexpected finding that the precursor effect was still present at 500 ms of temporal separation aroused our curiosity, and we decided to extend the interval durations even further, in the hope that the temporal range of the phenomenon could be delineated.

Methods

Subjects

Ten subjects from the same subject pool participated. Five had served in Experiment 2, three of them also in Experiment 1.

Stimuli

The same stimuli as in Experiment 2 were employed, but only the *Take the* and *Take this* precursors were used. Isolated test words were likewise omitted, to shorten the test and to avoid confusion in view of the rather long temporal separations between precursors and test stimuli. The four closure durations were 500, 1000, 1500, and 2000 ms. There were five blocks of 48 stimuli each. The procedure remained unchanged.

Results

The labeling functions are shown in the right-hand panel of Figure 2. A small boundary shift was evident that did not change much with closure duration. Even at a 2 s separation there was still a small average effect. The analysis of variance showed a significant main effect of condition [$F(1,9) = 9.02$, $p < .02$, $MSe = 14.70$], but no main effect of, or interaction with, closure duration.

FURTHER DISCUSSION OF EXPERIMENTS 1-3: THE TIME COURSE OF THE PRECURSOR EFFECT

To get a better overview of the time course of the precursor effect, a summary of the results of Experiments 1-3 is presented in Figure 3. The figure shows average category boundaries (in ms of VOT, determined by probit analysis of the average data) as a function of the temporal separation between precursor and test word (on a log scale).[2] The function on top shows that the VOT boundary in the *Take the* context decreased somewhat across the three experiments, from about 51 ms to 43 ms of VOT. It will be recalled that the *Take the* labeling function coincided with the function for isolated test words in Experiment 1 but began to separate from it in Experiment 2. Experiment 3 did not include isolated stimuli for comparison, but the trend appears to continue. The function in the center represents the *Take this* boundaries, and here there is a clear tendency for the VOT boundary to increase with closure duration, both between and within experiments, from about 27 ms to 39 ms of VOT. At the bottom of the figure, the difference between the *Take the* and *Take this* boundaries has been plotted, together with 95-percent confidence intervals.

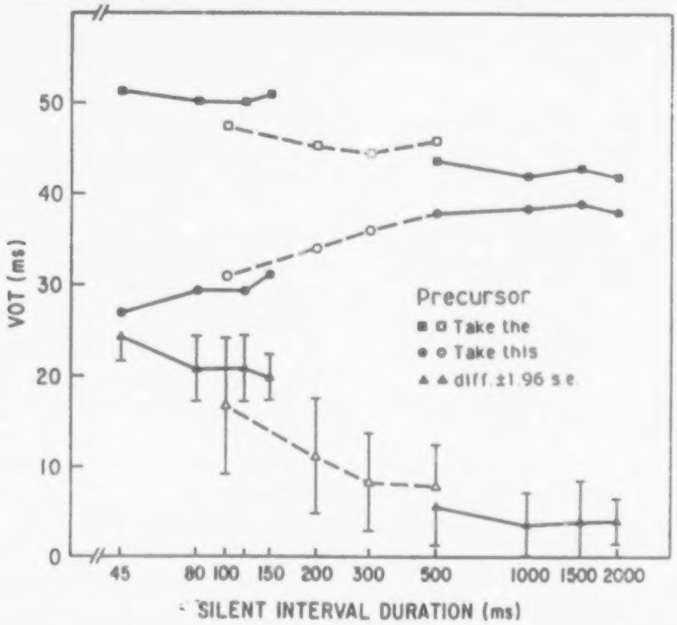


Figure 3. **Summary of results from Experiments 1-3: VOT boundary locations as a function of temporal separation between precursor and test word in the *Take the* and *Take this* precursor conditions. The function at the bottom shows the difference between the boundaries in the two precursor conditions, with 95% confidence intervals.**

The boundary difference can be seen to extend up to separations of 2 s and perhaps beyond. Considering how little change there was over the longer intervals, it might be speculated that there are really two effects, one decaying within about 1 s and the other extending over much longer intervals.

A puzzling feature of these data is that the two precursor conditions converge onto a category boundary that is some 10 ms shorter than that for isolated test words in Experiments 1 and 2. Either there was a bias to perceive voiceless stop consonants following any precursor, no matter how long the temporal separation, or some unknown factor introduced a bias to hear voiced stop consonants in isolated stimuli. This issue will be addressed in later experiments.[3]

Although the difference between the *Take this* and *Take the* precursor effects cannot be attributed to differences in speaking rate (if there were any unintended differences in rate, they were much too small to have any substantial effect), its decrease with increasing closure duration is remarkably similar to that of the speaking rate effect studied by Summerfield (1981; see his Figure 3). He found that the effect "was substantially reduced at an interval of 250 msec and reached an asymptote at an interval of about 1 sec" (p. 1082) and that it contained a "small long-lasting component" (p. 1082), the longest interval actually used being 8 s in his study. This time course is reminiscent of the decay times of the two components of auditory storage suggested by a large literature (reviewed by Cowan, 1984). Summerfield also showed that the speaking rate effect was primarily due to variations in the final syllable of the precursor (Summerfield, 1981: Exp. 4); in fact, the effect of selective rate variations in earlier syllables (more distant in time) was similar in size to the long-term component of the effect of overall rate variation. The speaking rate effect and the present precursor effect then perhaps have more in common than one might have thought: Both depend primarily on properties of the syllables or segments immediately preceding the closure, and both seem to be mediated by auditory memory for these properties. Still, it seems that they cannot be one and the same phenomenon. In the case of the speaking rate effect, the relevant acoustic property of the final segment is its duration or rate of spectral change. In the present case, it may be the presence versus absence of voicing.

EXPERIMENT 4: A VARIETY OF PRECURSORS

Experiment 4 tested this hypothesis by using a much wider variety of precursor phrases. Three of these phrases ended in a voiceless fricative or affricate, /s/, /č/, or /f/. The latter two consonants cannot form a syllable-initial cluster with /p/, so if ambiguity of word boundary location is a factor at all, it should not operate with them. Three other precursors ended in a voiced fricative or affricate, /z/, /ɟ/, or /v/. Because of the problem with final fricative devoicing in Experiment 2, these precursors were produced both in a natural and in a deliberately voiced version. Two additional precursors were included to test whether it was specifically fricative voicing/voicelessness or general presence versus absence of voicing that influenced VOT perception. These precursors ended in a vowel, /i/, and were produced either in a normal voice or in a whisper. If whispered precursors ending in a vowel had the same effect as precursors ending in a voiceless fricative, this would show the effect to be caused by general voicelessness of a preceding segment. To guard against artifacts due to token differences, multiple productions of each precursor were used. Moreover, each precursor was used in two different versions, one originally produced before /b/, the other before /p/. This served to test the influence of anticipatory coarticulatory information on voicing judgments. Such effects may have been present in our earlier experiments, even though they were in opposition to the precursor effect of primary interest.

Methods

Subjects

The 10 subjects of Experiment 3 were recalled for this experiment.

Stimuli

A new set of utterances was recorded by a different female speaker. She produced a variety of three-word phrases beginning with *Take* and ending with either *bins* or *pins*. The intervening words were: *six*, *these*, *such*, *large*, *tough*, *five*, and *three*.[4] The first six words form three pairs contrasting in the voicing of the final segment (/s/ vs. /z/, /č/ vs. /ɟ/, /f/ vs. /v/). A similar contrast was created for the precursor ending in /i/ by having the speaker produce the whole phrase either voiced or in a whisper. Each of the precursor phrases ending in a voiced fricative or affricate preceding *bins* was produced in an additional version, with the speaker (a graduate student of speech science) consciously trying to maintain voicing through the labial stop closure. Each of the phrases was produced five times in a random sequence. Thus there were a total of 95 utterances (8 precursors, each followed by either *bins* or *pins*, plus 3 deliberately voiced precursors followed by *bin§*, each produced five times). They were recorded in a sound-proof booth, sampled at 20 kHz, low-pass filtered at 9.6 kHz, and edited to remove the final *bins* or *pins* and the preceding stop closure. (Somewhat unintentionally, a significant amount of closure voicing was retained in tokens of *Take three*.)

One token each of *bins* and *pins* was used to construct a new VOT continuum, following the same procedures as in Experiment 1. The VOTs of the six stimuli used in the experiment were approximately 14, 23, 32, 40, 49, and 58 ms. Each of these stimuli was preceded by each precursor and also occurred five times in isolation, for a total of 600 stimuli, which were arranged in 5 blocks of 120, such that each unique stimulus occurred exactly once in each block. (That is, different blocks used different tokens of the same precursors.) There was a different random order within each block. The duration of the silent interval between precursor and test word (the stop closure) was fixed at 120 ms. Interstimulus intervals were 2 s, with longer intervals from time to time.

Procedure

The procedure remained the same as previously: The subjects identified the initial consonant of each test word as "B" or "P."

Results and Discussion

The results are summarized in Figure 4, which plots the average VOT boundaries for all precursor conditions. The left-most bar represents the boundary for isolated test words, against which the various precursor effects may be compared, with due caution.[5]

The first result to note is that all phonologically voiceless precursors (/s/, /č/, /f/, as well as whispered /i/) shifted the VOT boundary to shorter values, regardless of their original context. The shift following /s/ was somewhat larger than the others. Second, phonologically voiced precursors from the original *pins* context (/z(-p)/, /ɟ(-p)/, /v(-p)/) had exactly the same effect. Third, phonologically voiced precursors from the original *bins* context had very divergent effects, depending on the segment: /ɟ(-b)/ behaved like a voiceless segment, shifting the boundary down; /z(-b)/ and /v(-b)/ left the boundary unchanged, but deliberately voiced /z(-b)/ and /v(-b)/ caused a small upward shift; and /i(-b)/ caused a very large upward shift. In other words, there were large effects of anticipatory coarticulation in phonologically voiced precursor segments, except for /ɟ/, whereas phonologically voiceless precursor segments showed no coarticulation effects at all.

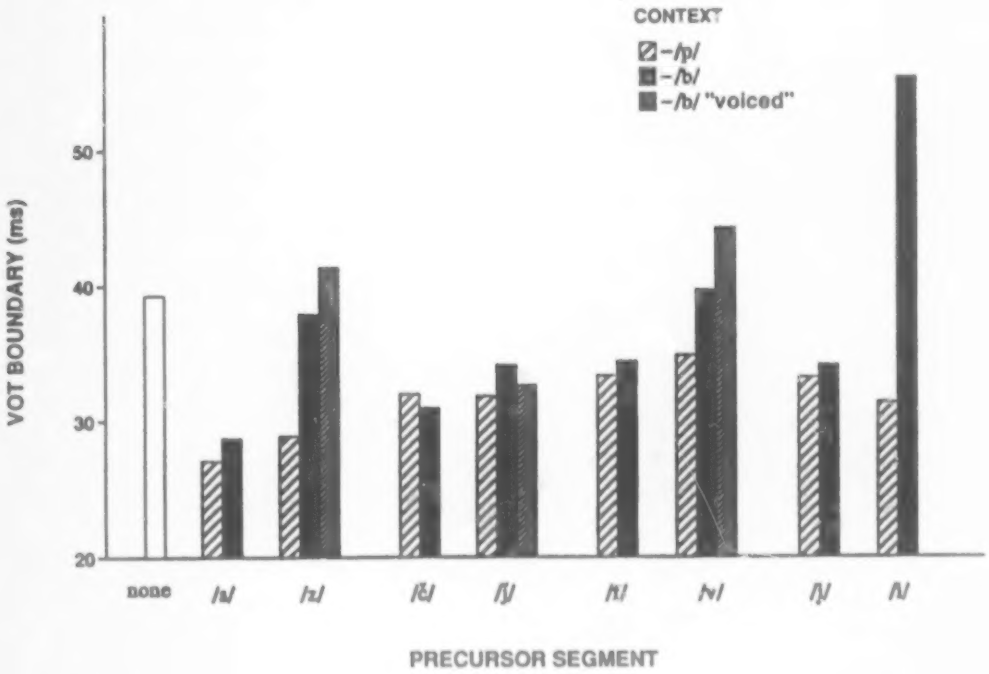


Figure 4. VOT boundaries as a function of final precursor segment and its original context in Experiment 4.

These patterns were shown to be statistically reliable in analyses of variance on the response percentages. In one analysis, isolated stimuli and the deliberately voiced precursor conditions were excluded, so that the remaining 16 conditions could be treated as representing three crossed factors pertaining to the final segments of the precursors: place of articulation (4 levels), phonological voicing status (2 levels), and original context (2 levels). All main effects and interactions were significant, so separate analyses were conducted on the effects of phonologically voiceless and voiced segments. For voiceless segments, the main effect of place of articulation reached significance [$F(3,27) = 3.38, p < 0.04$, MSe = 14.44], due to the shorter VOT boundaries following /s/, but there were no significant effects involving original context (i.e., coarticulation). For voiced segments, there were highly significant effects ($p < 0.0001$) of place of articulation, original context, and of their interaction, due to the abnormal behavior of /ǰ/ and the very large original context effect for /i/. Another analysis was conducted comparing only the voiced and deliberately voiced fricative precursors. The effect of deliberate voicing was significant [$F(1,9) = 5.94, p < 0.04$, MSe = 3.64], but in addition the main effect of place of articulation and the interaction were highly significant ($p < .0001$), again because /ǰ/ behaved like a voiceless fricative.

These results, in conjunction with acoustic analyses reported below, contribute considerably to our understanding of the precursor effects. First, they suggest that *any* voiceless segment preceding the stop closure causes a VOT boundary shift in favor of /p/ responses. This rules out the hypothesis (already called into question in Experiment 3) that the effect of a preceding /s/ was entirely due to the possibility of its forming a word-initial cluster with /p/. The effect of a preceding /s/, however, was somewhat larger than that of /č/, /f/, or whispered /i/. Thus there may have been some additional contribution of word boundary ambiguity, although the difference could also have been due to the specific acoustic properties of /s/. Primarily, however, the boundary shift seemed to be caused by the phonetic voicing status of the final segment of the precursor.

Precursors ending in a phonologically voiced fricative or affricate originally preceding /p/ had the same effect as phonologically voiceless precursors. The presumable reason for this was phonetic devoicing of the final segment. This was confirmed by an acoustic analysis of the precursors. The presence and strength of voicing during the final 128 ms of each precursor token were determined by measuring the amplitude of the spectral peak corresponding to the fundamental frequency (if any) in the discrete Fourier transforms of nine successive 25.6 ms waveform segments overlapping by 50%. The resulting curves of voicing amplitude as a function of time, averaged across the five tokens of each precursor, are shown in Figure 5 together with an approximate indication (based on waveform measurements) of where the fricative noise (or, in the case of /i/, the closure) began. It is evident that in /ǰ(-p)/, /z(-p)/, and /v(-p)/ (dotted lines) voicing ceased soon after the onset of frication. Moreover, the abnormal results for /ǰ/ preceding /b/ were obviously due to the fact that the speaker never maintained voicing through this affricate (Figure 5, upper left panel). In /z(-b)/ and /v(-b)/, however, voicing continued with diminishing strength through the fricative noise,[6] and in the deliberately voiced segments the amplitude of the fundamental was elevated further at the end. Clearly, therefore, the presence and degree of voicing in the final frication were responsible for the VOT boundary shifts following the various precursors. It is worth noting, though, that the most relevant acoustic parameter was apparently neither the voicing amplitude at precursor offset (in which case /z(-b)/ should have caused a larger upward shift of the VOT boundary than /v(-b)/, cf. Figure 4) nor, in the case of devoicing, the distance of voicing offset from the end of frication (in which case /v(-p)/ should have caused a smaller downward shift than /ǰ(-p)/ and /z(-p)/, not to mention /s/, /ʃ/, and /č/). Rather, the acoustic parameter most predictive of the perceptual results is the *slope of the decrease in voicing amplitude*: The more gradual the decrease, the more likely were subjects to perceive voiced stop consonants following the precursor.

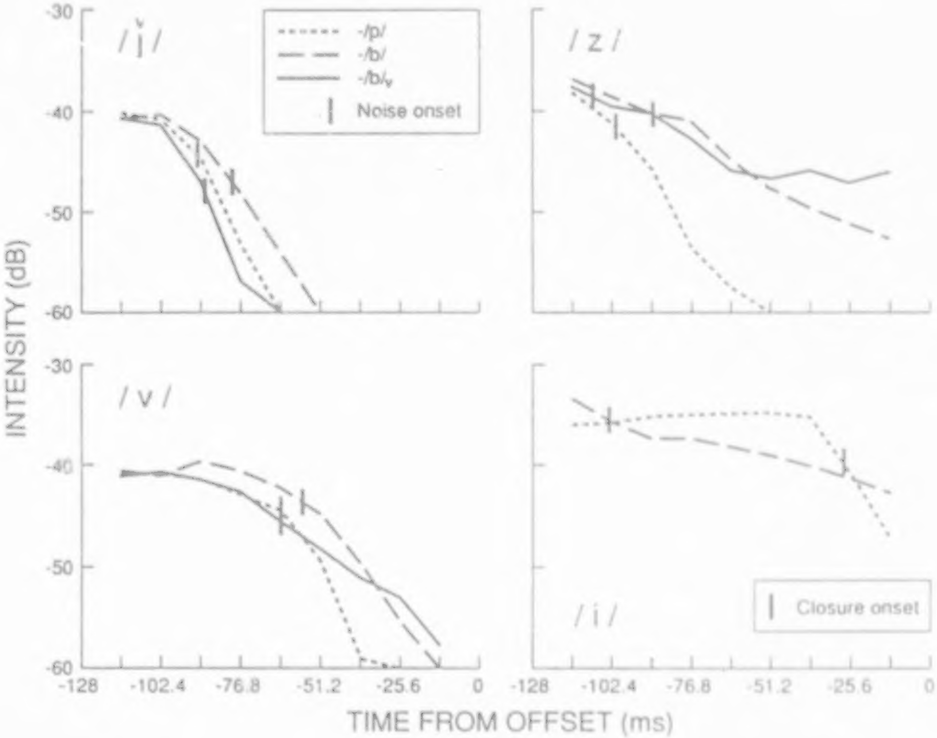


Figure 5. Intensity of the fundamental frequency during the final 128 ms of precursors ending in phonologically voiced segments, averaged over the five tokens of each type. Each data point derives from a window extending 12.8 ms to either side. The solid function represents the deliberately voiced condition. The vertical bars signify the average onset of aperiodic frication detectable in the waveform; for /i/, they indicate the beginning of the stop consonant closure.

This observation is further supported by the measurements for /i/ (Figure 5, lower right-hand panel). The perceptual effect of /i(-p)/, which equalled that of a voiceless fricative, seems paradoxical in view of the facts that voicing persisted right up to the end of the precursor (which included about 25 ms of closure voicing) and that the amplitude of its fundamental was higher than in voiced fricatives (and than in /i(-b)/, except at the very end). What /i(-p)/ had in common with the devoiced fricatives and affricates, however, was the rapid falloff in voicing amplitude, even though it occurred much later in the stimulus. The /i(-b)/ precursor, on the other hand, had a much more gradual decrease in voicing amplitude, which took place during the about 100 ms of voiced closure that happened to have been included, and which apparently caused the striking increase in voiced stop responses, despite the intervening 120 ms of silence. (That closure voicing at the beginning of a closure separating two stop consonants can influence the perceived voicing of the second stop consonant has been shown previously in Dutch by Berg, 1986.)

This identification of a *rate* parameter as a possible cause of the precursor effect raises the interesting possibility that the fairly large speaking rate effects observed by Summerfield (1981) were in fact partially mediated by the rate of change in voicing amplitude at precursor offset. The common finding is that voiced stop consonant responses to the critical VOT continuum increase as the precursor rate is slowed down. In Summerfield's experiments, changes in speaking rate were simulated by changing the time base of a speech synthesis program, which naturally also changed the amplitude curve associated with the fundamental at precursor offset (assuming that the amplitude of the final vowel was not constant). Diehl, Souther, and Convis (1980) used a similar method with precursors ending in /z/ and found only a very small effect; judging from their spectrograms, the voicing decay at precursor offset may not have been changed substantially. Miller, Green, and Schermer (1984) used natural speech precursors produced at different rates but kept the word *the* preceding the test words constant; they found only a very small rate effect. In an older, unpublished experiment, Summerfield (1975) varied only the steady-state portions of a synthetic precursor and also obtained only rather small rate effects on VOT perception. Kidd (1989), who used natural-speech precursors ending in *the*, found somewhat larger effects of precursor rate. However, the only truly large VOT boundary shifts (about 10 ms) contingent on variations in precursor rate were obtained by Summerfield (1981), whose precursors may have varied in the rate of voicing decay at offset. Therefore, this local phonetic factor may have contributed to the rate effect in that study.

At this stage in our research, the inevitable question arose of whether the observed phonetic effects are caused by acoustic stimulus properties alone, or whether they depend on the precursors being perceived as speech. The remaining experiments were intended to address this vexed issue which permeates contemporary speech perception research.

EXPERIMENT 5: NOISE PRECURSORS

It is notoriously difficult to dissociate genuinely speech-specific factors from purely auditory ones. A situation in which the same stimulus can be interpreted by listeners as either speech or nonspeech (cf. Best, Morrongiello, & Robson, 1981; Tomiak, Mullennix, & Sawusch, 1987) would be most suitable for this purpose but very difficult to contrive in the present precursor paradigm. We asked, instead, to what extent certain acoustic properties of the preceding speech segment—viz., its duration, intensity, and amplitude envelope—affect VOT perception at an auditory level. We investigated this by replacing the speech precursors with nonspeech noises sharing these properties but having no spectral structure (cf. Gordon, 1988). Since all precursors here were aperiodic noises, we expected that all of them would cause a downward VOT boundary shift relative to a no-precursor baseline, and that any differences among precursors in the magnitude of these shifts would reveal the role of the acoustic properties mentioned. This experiment, then, investigated the possible role of factors *other than* the presence of voicing near the end of the precursor.

Methods

Subjects

Nine new subjects from the same general pool participated.

Stimuli

Stimuli, test arrangement, and procedure were exactly the same as in Experiment 4, except that each speech precursor was replaced with amplitude-modulated broadband noise. This noise was generated from each individual speech precursor by randomly reversing the polarity of sampling points with a probability of 0.5 (Schroeder, 1968). Thus each noise had exactly the same duration, overall intensity, and amplitude envelope as the speech it was derived from, but it lacked spectral structure and was not identifiable as speech.[7]

Results and Discussion

The results are easily summarized. The average VOT boundaries in all 19 precursor conditions as well as for isolated stimuli were within 3 ms of each other, and the analysis of variance revealed no significant differences whatsoever. Thus the noise precursors did not affect the VOT boundary at all.

One possible explanation of this result is that the listeners dissociated the nonspeech precursors from the following speech, attributing them to a different sound source (cf. Gordon, 1988). This interpretation would negate relatively peripheral auditory interactions as an explanation of the earlier VOT boundary shifts. An alternative explanation, however, is that the broadband noise precursors lacked precisely the acoustic property that was responsible for the speech precursor effects—viz., periodicity. Although we attributed the downward shift of the VOT boundary caused by precursors ending in voiceless fricatives (Exp. 4) to the lack of periodicity at precursor offset, perhaps this effect was caused by the early and abrupt cessation of voicing rather than by its absence later on. If so, precursors totally devoid of periodicity would naturally be ineffective. However, this explanation is in conflict with the finding in Experiment 4 that the effect of whispered precursors was similar to that of precursors ending in a voiceless fricative.

A third possibility is that precursors ending in voiceless segments simply have no effect at all, that such an effect was wrongly suggested by the unstable boundary for isolated test stimuli, and that all true precursor effects are upward shifts of the VOT boundary due to the presence of voicing at precursor offset. This interpretation, however, is at odds with the results of Experiments 1-3, which showed a decrease over time in the effect of a precursor ending with /s/.

A final explanation is that the noise precursors in this experiment did have an effect, but that the boundary for isolated stimuli was at too short a VOT value. Indeed, the VOT boundary for isolated test words was at 29 ms of VOT, 10 ms shorter than the boundary for the very same stimuli in Experiment 4, and in fact close to the /s/ precursor boundaries in that experiment. Had the boundary for isolated stimuli been in the same place as in Experiment 4, we could have concluded that the preceding noise caused a boundary shift, just like the voiceless fricative precursors. It is becoming increasingly clear that the general experimental context has a substantial effect on the VOT boundary for isolated test words, and that comparisons with this baseline are therefore problematic. Apparently, the presence of nonspeech noise precursors in the present test caused the baseline boundary to move towards a much shorter VOT.

EXPERIMENT 6: SPEECH AND NOISE PRECURSORS

Experiment 6 attempted to circumvent this problem by including both speech precursors ending in voiceless fricatives and nonspeech precursors in the same test. It was expected that this manipulation would place the baseline boundary between the locations observed in Experiments 4 and 5. The question was whether the boundaries in all precursor conditions would be at shorter VOTs.

Methods

Subjects

Eleven subjects participated, several of whom had also served in earlier experiments.

Stimuli

In addition to isolated test words, the *Take six*, *Take such*, and *Take tough* precursors of Experiment 4 (from the *bins* context only) were used, as well as their amplitude-modulated noise versions from Experiment 5. Six precursor conditions times six test stimuli plus isolated test words resulted in 42 stimuli altogether, which were presented five times in random order, with different tokens in each block. The procedures were the same as previously.

Results and Discussion

Our expectation concerning the boundary for isolated stimuli was indeed confirmed: The boundary was at about 33 ms of VOT, almost halfway between the boundaries observed in Experiments 4 and 5. However, the category boundaries in all precursor conditions fell within 2 ms of each other and did not differ significantly from each other. This was surprising in view of the differences among the three speech precursors obtained in Experiment 4. Even more surprising, however, was the fact that there was absolutely no difference between the precursor and baseline boundaries.

To highlight the contrast with the previous findings, the results of Experiment 6 are plotted in Figure 6 together with the corresponding data from Experiments 4 and 5. The uniformly longer boundaries in Experiment 6 than in Experiment 5 could reflect simply the difference in subject groups; however, the ineffectiveness of the speech precursors in Experiment 6 relative to Experiment 4 is puzzling. One possible explanation of these results is that the presence of nonspeech precursors in the test led subjects to employ an attentional strategy of dissociating *all* precursors from their following context. Such a voluntary segregation of components of a speech stream seems not very plausible. Alternatively, if the absence of differences among the speech precursors is disregarded, it is still possible that the boundary for isolated stimuli was at "too short" a VOT in Experiment 6, and that all precursors "really" caused a downward shift in the boundary. A final attempt to circumvent this baseline problem was undertaken in Experiment 7, by employing a separate test containing only isolated test stimuli.

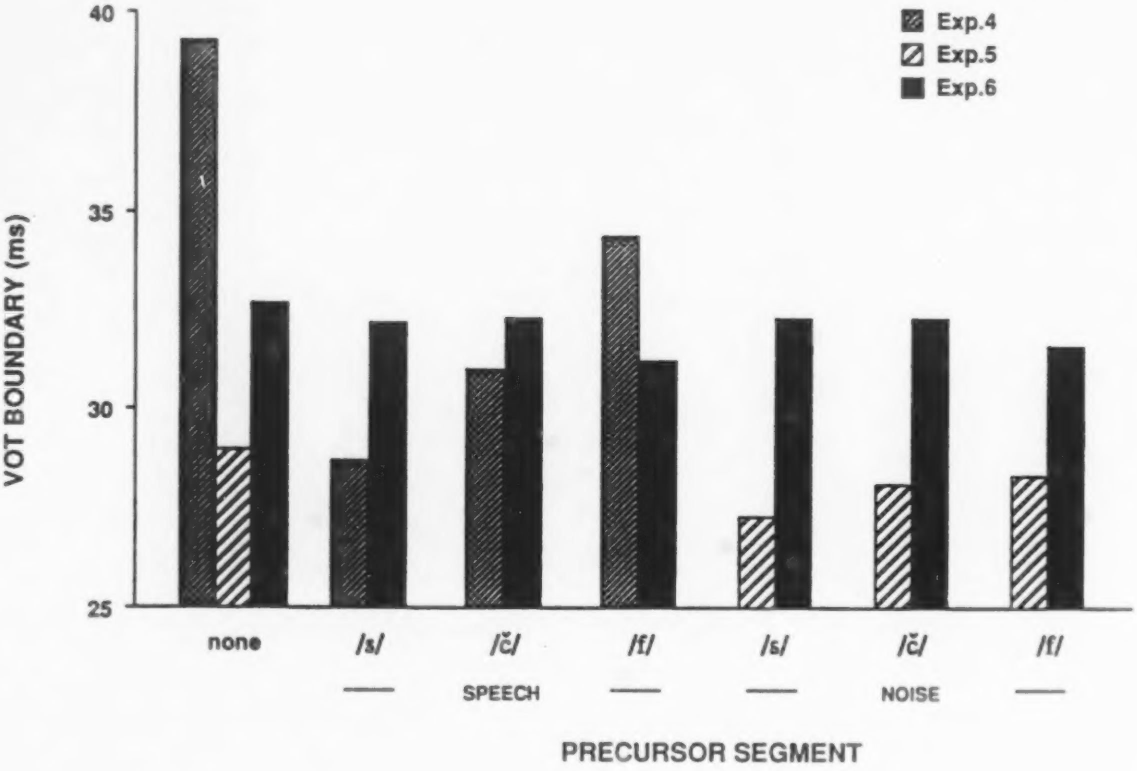


Figure 6. **Results of Experiment 6 in relation to selected results from Experiments 4 and 5: VOT boundaries as a function of precursor condition.**

EXPERIMENT 7: FILTERED SPEECH PRECURSORS

Experiment 7 had two purposes: First, to obtain an independent baseline estimate and to assess directly the influence of various precursor environments on the baseline boundary itself; second, to examine the role of low-frequency energy in the precursor effect. By focusing on the auditory correlate of voicing, this experiment was complementary to Experiments 5 and 6. Experiment 7 used precursors with final voiced segments as well as precursors with final voiceless segments, so a precursor effect could be established by comparing these two types, without reference to a baseline (as in Experiments 1-4).

The experiment included a separate baseline test and four conditions with precursors, each also including isolated stimuli. The first condition partially replicated Experiment 4; the precursors chosen were *Take five* and *Take these*, from either the original *bins* or the *pins* context. Thus the comparison was between phonetically voiced and voiceless versions of the same phonological structures, which had yielded significantly different boundaries in Experiment 4. These results were expected to be replicated here; the question was where the boundary for the embedded isolated stimuli would be located relative to that obtained in the separate baseline test.

In the second condition, the precursor phrases were high-pass filtered (HPF) to remove the fundamental frequency and its lowest harmonics. The result was intelligible speech of telephone-like quality. Although, as is well known, physical absence of the fundamental frequency does not prevent its perception in a complex periodic sound (Schouten, 1970), this condition served to investigate whether low-frequency energy as such is involved in the precursor effect. If it is, the difference between phonetically voiced and voiceless precursors should be reduced or eliminated by high-pass filtering.

In the third condition, the precursor phrases were low-pass filtered (LPF), so that practically only the fundamental frequency remained. The precursors retained some of the time and amplitude structure of the originals, but they were barely intelligible as speech. If low-frequency energy as such is important, particularly its amplitude contour at precursor offset, then these precursors should still produce effects similar to the originals.

The fourth condition was intended to investigate the possibility that low-frequency voicing information is effective only when it occurs in the context of amplitude-modulated high-frequency energy, as in speech. To that end, new precursors were constructed by combining the LPF stimuli of the third condition with nonspeech noises derived from the same original precursor phrases (LPF+N). These noises (similar to those used in Experiment 5) were first high-pass filtered to prevent auditory masking of the low-frequency information by the noise. We thought that addition of this telephone-quality noise to the voicing might enhance the effectiveness of the voicing information by providing a more precise definition of voice offset time, and by making the stimulus somewhat more speechlike (though not quite intelligible). On the other hand, we realized that, to the extent that there is an independent effect of voicelessness in the precursor, addition of the noise might cause a boundary shift to shorter VOTs.

All four conditions included isolated *bins/pins* stimuli to further examine shifts in their boundary across conditions. The hypothesis suggested by Experiments 4-6 was that the overall relative amount of voicing in the stimulus ensemble affects the baseline boundary. If so, the boundary should be at the longest value in Condition 1 and at the shortest in Condition 4, with the results for Conditions 2 and 3 depending on whether low-frequency energy as such is important.

Methods

Subjects

Twelve new subjects participated.

Stimuli

In Condition 1 (intact speech precursors), the carrier phrases used were *Take five* and *Take these* from the *bins* and *pins* contexts.[8] For Condition 2 (HPF precursors), the precursor phrases were digitally filtered with a lower cutoff frequency of 1000 Hz and stop-band attenuation of 50 dB, which insured a steep slope. For Condition 3 (LPF precursors), the upper cutoff frequency was set at 500 Hz, and the stop-band attenuation was 50 dB. For Condition 4 (LPF+N precursors), amplitude-modulated noise versions of the precursors were prepared as in Experiment 5; subsequently, they were high-pass filtered (like the speech precursors for Condition 2) and digitally added to the LPF precursors of Condition 3.

Procedure

The baseline sequence of isolated test words was presented first, followed by the four experimental conditions and, finally, a repetition of the baseline sequence. The order of the four conditions was counterbalanced across subjects. Otherwise, the procedure was the same as previously.

Results and Discussion

The results are shown in Figure 7 in terms of average VOT boundaries. Consider first the results for isolated stimuli. The boundaries in the pre- and posttest were at 23.1 and 25.6 ms of VOT, respectively. The difference was nonsignificant [F(1,11) 3.47, $p < .09$]. These boundaries are shorter than in any preceding experiment, and in fact are close to what one should expect for stimuli along a bilabial VOT continuum. This finding alone suggests that all previous "baseline" boundaries were artificially elevated due to some influence of general test environment.

Indeed, the boundaries for the isolated stimuli embedded in the present precursor test sequences were at longer VOTs also, with the longest boundary in Condition 1 (intact precursors). A one-way ANOVA across the six VOT boundaries for isolated stimuli showed a highly significant difference [F(5,55) 7.77, $p < .0001$, MSe = 20.54]. Subsequent Newman-Keuls tests showed the boundary in the intact condition to be significantly ($p < .01$) longer than the pre- and posttest boundaries, and also longer than the boundary in the LPF condition. Furthermore, the HPF and LPF+N boundaries were significantly longer than the pretest boundary. Only the LPF boundary did not differ significantly from the pre- and posttest boundaries. Essentially, each boundary fell into the range of the boundaries following the precursors in that condition (see Figure 7).

With regard to the precursor conditions, it is evident from Figure 7 that only the intact condition generated a substantial precursor effect.

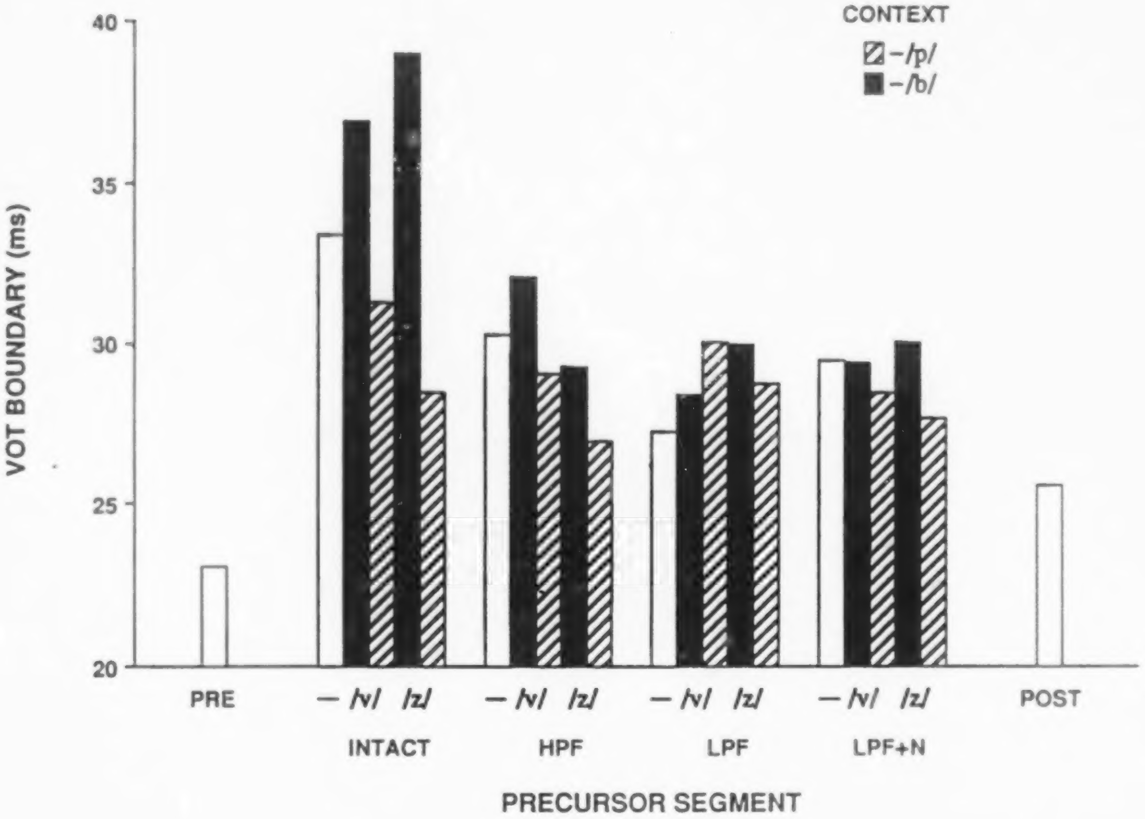


Figure 7. VOT boundaries as a function of various precursor conditions in Experiment 7.

A 2 × 2 ANOVA (Voicing by Place of Articulation of the precursor-final segment) on these data revealed a significant effect of Voicing [F(1,11) 20.24, $p < .001$, MSe = 38.39], with longer VOT boundaries following voiced than voiceless segments, and a significant Voicing by Place interaction [F(1,11) 17.41, $p < .002$, MSe = 4.26], due to the voicing effect being larger for /z/ than for /v/. These data are in good agreement with those of Experiment 4 (Figure 4). The effect of precursor Voicing was much smaller in the HPF condition, but it was still reliable [F(1,11) 6.51, $p < .03$, MSe = 12.58]. Instead of an interaction, there was a main effect of Place [F(1,11) 7.40, $p < .02$, MSe = 9.50], due to longer boundaries following /v/ than following /z/. All HPF boundaries were in the same range as the boundaries following voiceless intact precursors. The boundaries in the LPF and LPF+N conditions were in the same range also. There were no significant LPF precursor effects. A small effect of Voicing emerged again in the LPF+N condition, however [F(1,11) 5.72, $p < .04$, MSe = 5.83].

In summary, these data show that only the voicing information in intact speech precursors has a substantial effect on the VOT boundary; the presence of low-frequency energy in the precursor seems to be necessary (though a small effect was obtained with HPF precursors) but not sufficient. Since high-pass filtering merely degrades phonetic voicing information, whereas low-pass filtering largely eliminates it and the addition of noise restores it only minimally, the results suggest that it is the phonetic voicing information in the precursor-final segment, rather than the simple auditory property of low-frequency energy, that causes the precursor effects.

The results also demonstrate convincingly that the boundary for isolated stimuli is unstable and shifts into the range of boundaries caused by the precursors in any particular condition. This shift is invariably towards longer VOTs.

GENERAL DISCUSSION

The present series of experiments demonstrated two effects of preceding phonetic context on the perception of stop consonant voicing: an effect of the immediately preceding phonetic segment, and an effect of the whole stimulus ensemble in the test sequence.

The effect of the immediately preceding segment is assimilative in nature and appears to derive from the segment's phonetic voicing status. Experiments 2 and 4 showed clearly that the phonological voicing category of the precursor segment is irrelevant: When voiced fricatives are devoiced, they exert the same precursor effect as voiceless ones. Also largely irrelevant is the closure duration intervening between precursor segment and test word (Experiment 1; see also McCasland, 1977) and the capability of the precursor segment to form a tautosyllabic cluster with the initial consonant of the test word (Experiment 4). Experiment 4 suggested that the most relevant phonetic property may be the degree of abruptness of voicing cessation in the precursor, and Experiment 7 indicated that only intact speech produces a sizeable precursor effect. Perhaps the most interesting finding was the relatively slow temporal decay of the effect (Experiments 2 and 3): Although it decreased in size over the first 300-500 ms, a residual precursor effect extended over several seconds of temporal separation.

These findings imply that listeners base their decision about the voicing status of an ambiguous word-initial stop consonant on a wide temporal interval that includes the segment (or its offset) preceding the stop closure. Such a temporal integration account of voicing perception was elaborated by Chasaide (1985) and also hinted at by Berg (1986). Acoustic cues to the voicing of an intervocalic stop are commonly present before its closure begins, certainly within words (Lisker, 1986), but probably also across word boundaries. The coarticulatory effects observed in Experiment 4 attest to this. The precursor effect to be explained, however, does not rest on proper cues to stop consonant voicing; rather, it reflects voicing properties of the pre-closure segment that are only partially conditioned by the following stop consonant. In Repp's (1982) classification, it is more like a context effect than like a cue integration effect (or trading relation).

Such context effects are usually explained in one or both of two ways (see Repp, 1982; Diehl & Kluender, 1989): They may reflect a perceptual compensation for certain articulatory/acoustic regularities in speech production, or they may be due to a purely auditory processing interaction. Since our knowledge of the processing of complex auditory signals is still quite limited (see, e.g., Yost & Watson, 1987), auditory interactions can never be ruled out with certainty. Experiment 5, at least, showed that variations in the precursor amplitude envelope are irrelevant, and Experiment 7 showed little effect of low-frequency energy as such; also, the precursor effects obtained over long temporal separations (Experiments 2 and 3) are most likely not due to auditory processing interactions. They do reflect memory for the precursor, however, and we have noted earlier that the temporal decay of the precursor effect is consistent with the two components of auditory memory discussed by Cowan (1984). Clearly, the precursor information that causes the VOT boundary shift is not retained in categorical form, otherwise temporal separation would not have reduced the precursor effect substantially. The memory involved, however, could be for articulatory gestures and states rather than for complex auditory properties, and the dynamic and static components of such an articulatory memory might correspond to the two time constants of the decay function.

If an explanation of the precursor effect were to be found in a regularity of speech production that listeners have tacit knowledge of, that regularity would have to be the devoicing of final voiced consonants when they precede a word beginning with a voiceless consonant. Although such devoicing may occur even preceding a voiced consonant (cf. Experiments 2 and 4), it is not the rule in that context (Haggard, 1978; Ladefoged, 1975). Thus the occurrence of a voiced fricative at the end of a precursor phrase is a clear indication that the test word begins with a voiced consonant. A voiceless fricative, on the other hand, is not predictive of the voicing of following segment. These facts suggest that a context effect should be exerted by voiced precursor segments, but not by voiceless ones. The present results, however, suggest that both have an effect (see Figures 3 and 7). Thus they provide no strong evidence that the precursor effects are specifically linked to the phenomenon of final fricative devoicing.

The results of Experiment 6, in which both speech and nonspeech precursors were employed, remain somewhat puzzling. Not only were the effects of the two types of precursors equivalent, but differences among the effects of the speech precursors obtained in Experiment 4 were no longer present. It is probably most parsimonious to consider these results merely as a failure to replicate these differences among precursors, rather than as evidence that the subjects, under the influence of the nonspeech precursors, managed to disengage the speech precursors perceptually from the following test words. Previous related research (Repp, 1975) suggests that phonetic context effects are highly resistant to attentional manipulation. On the other hand, it might be noted that the senior author, who served as a pilot subject for some of the experiments, never showed any effects of preceding context on his VOT boundary. There were considerable variations in the magnitude of precursor effects for individual subjects in the experiments, but there was hardly ever a subject who did not show any effect at all. These observations suggest that the precursor effect may depend on subjects' (in)ability to listen analytically to speech.

The present series of experiments revealed a second kind of context effect, due to the test environment, which was unexpected and was initially attributed to the stimuli varying in VOT. Only in Experiment 7 it became clear that the test stimuli were not abnormal at all and had a VOT boundary at the value expected for labial stop consonants (about 25 ms) when presented by themselves in a separate test. Whenever they were interspersed among stimuli preceded by precursor phrases, their VOT boundary was shifted dramatically in favor of voiced responses. The magnitude of this shift varied across experiments, and some of this variation surely must be attributed to variability among subject groups. Nevertheless, there was clearly a relation to the nature of the precursors employed in a test: When all precursors were intact speech (Experiments 1-4), the VOT boundary for the "baseline" stimuli was always above 35 ms; this was also true for the asymptotic boundary of stimuli separated from speech precursors by long silent intervals (Experiment 3). Only in the intact precursor condition of Experiment 7 there was a somewhat shorter boundary, at 33 ms. A boundary at that value was also obtained when the test contained a mixture of speech and nonspeech precursors (Experiment 6), whereas boundaries of 30 ms or less were obtained in conditions employing only nonspeech precursors (Experiment 5) or filtered speech precursors (Experiment 7). Even in those conditions, the boundary still seemed somewhat elevated with respect to that for isolated stimuli in a separate test.

The mechanism causing these boundary shifts remains obscure. Although in Experiments 1-3 the "baseline" boundary was at a rather high value, the data of Experiments 4-7 are consistent with the hypothesis that the baseline boundary fell in the midrange of the other boundaries obtained in the experiment. Thus the phenomenon could be related to that of adaptation level (Helson, 1964): Subjects were not able to maintain separate perceptual criteria for test stimuli presented in and out of context, and they treated the baseline stimuli (which were relatively infrequent in each test) as if they had been preceded by a precursor—

perhaps the one that occurred last. It might be noted here that, although the present test stimuli had a "normal" VOT boundary when presented in a separate test, they did contain residual cues favoring the "voiced" category (see Footnote 1), which may have assumed a more prominent role in the precursor test environment. A clearer elucidation of the causes of this phenomenon will require further research, though it is not clear whether anything important about speech perception will be learned thereby. The finding is of methodological importance, however, as it draws attention to the problem of finding an appropriate "baseline" condition in studies of context effects.

REFERENCES

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: Fo shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 25-33). New York: Academic Press.

Berg, R. J. H. van den (1986). The effect of varying voice and noise parameters on the perception of voicing in Dutch two-obstruent sequences. *Speech Communication, 5*, 355-367.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*, 191-211.

Chasaide, A. Ni (1985). *Preaspiration in phonological stop contrasts.* Doctoral dissertation, University College of North Wales, Bangor.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*, 341-370.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America, 74*, 1126-1135.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America, 27*, 761-764.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121-144.

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27*, 435-443.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*, 137-146.

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics, 6*, 95-102.

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 736-748.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich, Inc.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language, 33*, 42-49.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29*, 3-11.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.

McCasland, G. P. (1977). English stops after /s/ at medial word-boundary. *Phonetica, 34*, 218-228.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36*, 329-337.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61*, 1352-1361.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Repp, B. H. (1985). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 799-813.

Repp, B. H., & Lin, H.-B. (1989). Effects of preceding context on discrimination of voice onset times. *Perception & Psychophysics, 45*, 323-332.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44*, 1735-1736.

Schouten, J. F. (1970). The residue revisited. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 41-54). Leiden: Sijthoff.

Summerfield, Q. (1975). Cues, contexts, and complications in the perception of voicing contrasts. *Speech Perception* (Department of Psychology, The Queen's University of Belfast), 2, No. 4, 99-129.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1074-1095.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81*, 755-764.

Wingate, A. H. (1982). A phonetic answer to a phonological question. *UCLA Working Papers in Phonetics, 54*, 11-27.

Yost, W. A., & Watson, C. S. (Eds.) (1987). *Auditory processing of complex sounds*. Hillsdale, NJ: Lawrence Erlbaum Associates.

FOOTNOTES

[*] *Journal of Experimental Psychology: Human Perception and Performance*, in press.

[†] Also Department of Linguistics, University of Connecticut, Storrs. Now at Speech and Hearing Sciences, The Graduate Center, CUNY, New York.

[1] The average category boundary for isolated stimuli was at the unusually long value of 52 ms of VOT, as compared to 35 ms previously. The reason for this large difference is not clear, since the range of VOTs was merely narrowed, not shifted, with respect to the earlier experiment. The category boundary was at a relatively long VOT for a bilabial continuum even in the earlier experiment. One reason for this bias may lie in the method of stimulus construction, which retained the acoustic properties of bin at, and following, the onset of voicing. These properties included the amplitude and fundamental frequency contours, which are known to be secondary voicing cues (see Darwin & Seton, 1983; Abramson and Lisker, 1985). The real reason, however, became apparent in later experiments of this series (see below).

[2] The individual results of those subjects who participated in two or all three of the experiments are consistent with the pattern shown in Figure 3.

[3] A listener might expect isolated words to be produced at a slower rate than phrase-final words, so that the same stimuli

might sound relatively faster in isolation than following a precursor. If so, the boundary for isolated stimuli should be at a shorter VOT (cf. Summerfield, 1981), which is the opposite of what we found. We also note that, although the boundary obtained for isolated stimuli in our pilot study (Figure 1) is closer to the asymptotic boundary following precursors in Experiment 3 (Figure 3), peak performance in a discrimination task employing the very same isolated stimuli (Repp & Lin, 1989) occurred at an even shorter VOT. Such a dissociation of labeling and discrimination performance is unusual in the categorical perception literature and suggests that the boundary for isolated stimuli was influenced by as yet undetermined factors.

[4] Meaningful phrases were chosen to increase the naturalness of the materials, at the cost of relinquishing some control over their phonetic composition. It was assumed that the segments preceding the final consonant of the precursor would not exert any effect on VOT perception.

[5] It should be noted that the boundary for isolated stimuli in this experiment was at a much shorter VOT value than in Experiments 1 and 2, closer to that obtained in the pilot study (Figure 1) and also consistent with the asymptotic boundary in the precursor conditions of Experiment 3 (Figure 3). See also Footnotes 1 and 3, and later experiments.

[6] The data for /z(-b)/ are not representative of individual tokens: Two were devoiced, like /z(-p)/, and three were more strongly voiced, like /z(-b)/.

[7] Following the random sample-reversal procedure, the noise precursors had a flat spectrum, like white noise. However, because they were converted into sound at the same time as the digitized test words, which had been input with high-frequency pre-emphasis, they underwent high-frequency de-emphasis at the D/A conversion stage. Thus their actual spectrum had a falling slope (about 6 dB per octave above 1000 Hz, and less below).

[8] The deliberately voiced versions (see Experiment 4) were not used because they differed somewhat in intonation and overall amplitude from the "natural" versions, whose contextual variants sounded very similar except for the presence or absence of final fricative devoicing.

Appendix

SR #	Report Date	DTIC #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-90
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329/AS	
SR-97/98	January-July 1989	PB 90-121161/AS	
SR-99/100	July-December 1989	PB 90-226143/AS	

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Haskins Laboratories Status Report on
Speech Research

SR-101/102
JANUARY-JUNE 1990

Contents

- ○ The Alphabetic Principle and Learning to Read
  Isabelle Y. Liberman, Donald Shankweiler, and Alvin M. Liberman 1
- ○ Language Development from an Evolutionary Perspective
  Michael Studdert-Kennedy 14
- ○ Phonemic Segmentation as Epiphenomenon: Evidence from the History of Alphabetic Writing
  Alice Faber 28
- ○ Phonemic Similarity Effects and Prelexical Phonology
  Georgije Lukatela and M. T. Turvey 41
- ○ Automatic and Prelexical Computation of Phonology in Visual Word Identification
  Georgije Lukatela and M. T. Turvey 74
- ○ Phonological Ambiguity and Lexical Ambiguity: Effects on Visual and Auditory Word Recognition
  Ram Frost, Laurie B. Feldman, and Leonard Katz 85
- ○ Morphological Relationships Revealed Through the Repetition Priming Task
  Laurie Beth Feldman 101
- ○ Listener-talker Attunements in Speech
  Carol A. Fowler 110
- ○ An Electropalatographic and Acoustic Study of Consonant-to-Vowel Coarticulation
  Daniel Recasens 130
- ○ Coarticulation Is Largely Planned
  D. H. Whalen 149
- ○ The Perceptual Effects of Child-Adult Differences in Fricative-Vowel Coarticulation
  Susan Nittrouer and D. H. Whalen 177
- ○ "Targetless" Schwa: An Articulatory Analysis
  Catherine P. Browman and Louis Goldstein 194
- ○ Methods for Least-squares Parameter Identification for Articulatory Movement and the Program PARFIT
  Richard S. McGowan, Caroline L. Smith, Catherine P. Browman, and Bruce A. Kay 220
- ○ "Clock" and "Motor" Components in Absolute Coordination of Rhythmic Movements
  M. T. Turvey, R. C. Schmidt, and Lawrence D. Rosenblum 231
- ○ An Analogy between the Mucosal Waves of the Vocal Folds and Wind Waves on Water
  Richard S. McGowan 243
- ○ Laryngeal Control in the Plosives of Standard Thai
  Arthur S. Abramson 250
- ○ Distinctive Vowel Length: Duration vs. Spectrum in Thai
  Arthur S. Abramson and Nianqi Ren 256
- ○ Effects of Preceding Context on the Voice-Onset-Time Category Boundary
  Bruno H. Repp and Hwei-Bing Lin 269

Appendix 289